

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



(43) Date de la publication internationale
30 novembre 2000 (30.11.2000)

PCT

(10) Numéro de publication internationale
WO 00/71710 A2

(51) Classification internationale des brevets⁷: C12N 15/12,
C12Q 1/68, C07K 14/47

CLEPET, Christian [FR/FR]; 8 Rue Charles Fournier,
F-75013 Paris (FR).

(21) Numéro de la demande internationale:
PCT/FR00/01426

(74) Mandataire: BOUVET, Philippe; Aventis Pharma S.A.,
Direction Brevets, 20 Avenue Raymond Aron, F-92165
Antony Cedex (FR).

(22) Date de dépôt international: 25 mai 2000 (25.05.2000)

(25) Langue de dépôt: français

(26) Langue de publication: français

(30) Données relatives à la priorité:
99/06587 25 mai 1999 (25.05.1999) FR
60/139,450 16 juin 1999 (16.06.1999) US

(81) États désignés (*national*): AE, AG, AL, AM, AT, AU, AZ,
BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK,
DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID,
IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT,
LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ,
PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT,
TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) États désignés (*régional*): brevet ARIPO (GH, GM, KE,
LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), brevet eurasien
(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen
(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU,
MC, NL, PT, SE), brevet OAPI (BF, BJ, CF, CG, CI, CM,
GA, GN, GW, ML, MR, NE, SN, TD, TG).

(71) Déposant (*pour tous les États désignés sauf US*): AVEN-
TIS PHARMA S.A. [FR/FR]; 20 Avenue Raymond Aron,
F-92160 Antony (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (*pour US seulement*): DENEFLÉ,
Patrice [FR/FR]; 45 Avenue des Fusillés de Châteaubri-
ant, F-94100 Saint Maur (FR). ROSIER-MONTUS,
Marie-Françoise [FR/FR]; 21 Rue des Baconnets,
F-92160 Antony (FR). ARNOULD-REGUIGNE, Is-
abelle [FR/FR]; 112 Rue de Bry, F-94430 Chennevières
sur Marne (FR). PRADES, Catherine [FR/FR]; 30
Avenue du Général de Gaulle, F-94320 Thiais (FR).

Publiée:

— Sans rapport de recherche internationale, sera republiée
dès réception de ce rapport.

En ce qui concerne les codes à deux lettres et autres abrégia-
tions, se référer aux "Notes explicatives relatives aux codes et
abréviations" figurant au début de chaque numéro ordinaire de
la Gazette du PCT.

(54) Title: EXPRESSION PRODUCTS OF GENES INVOLVED IN DISEASES RELATED TO CHOLESTEROL METABOLISM

(54) Titre: PRODUITS D'EXPRESSION DE GENES IMPLIQUES DANS DES AFFECTIONS ASSOCIEES AU METABO-
LISME DU CHOLESTEROL

(57) Abstract: The invention concerns nucleic acids expressed from genes located in the human genome in the 9q31-34 region of chromosome 9, likely to be involved in diseases genetically related to said chromosomal locus, in particular diseases of the plasmatic lipoprotein metabolism, more particularly the reverse transport of cholesterol. The invention also concerns polypeptides encoded by certain nucleic acids and antibodies specifically directed against such polypeptides, useful as diagnostic reagents. The invention further concerns vectors and recombinant host cells comprising said nucleic acids or fragments thereof.

(57) Abrégé: La présente invention concerne des acides nucléiques exprimés à partir de gènes localisés dans le génome humain dans la région 9q31-34 du chromosome 9, susceptibles d'être impliqués dans des maladies génétiquement liées à ce locus chromosomique, notamment les affections du métabolisme des lipoprotéines plasmatiques, plus particulièrement le transport inverse du cholestérol. L'invention concerne également des polypeptides codés par certains de ces acides nucléiques ainsi que des anticorps dirigés spécifiquement contre de tels polypeptides, utiles comme réactifs de diagnostic. L'invention est enfin relative à des vecteurs et des cellules hôtes recombinantes comprenant ces acides nucléiques ou des fragments de ceux-ci.

WO 00/71710 A2

**PRODUITS D'EXPRESSION DE GENES IMPLIQUES DANS DES AFFECTIONS
DU METABOLISME DU CHOLESTEROL**

La présente invention concerne des acides nucléiques exprimés à
5 partir de gènes localisés dans le génome humain dans la région 9q31-
34 du chromosome 9, susceptibles d'être impliqués dans des maladies
génétiquement liées à ce locus chromosomique, notamment les
affections du métabolisme des lipoprotéines plasmatiques, plus
particulièrement le transport inverse du cholestérol.

10 L'invention concerne également des polypeptides codés par
certains de ces acides nucléiques ainsi que des anticorps dirigés
spécifiquement contre de tels polypeptides, utiles comme réactifs de
diagnostic.

L'invention est enfin relative à des vecteurs et des cellules hôtes
15 recombinantes comprenant ces acides nucléiques ou des fragments de
ceux-ci.

Les lipoprotéines, complexes protéiques incorporant des lipides
permettant le transport de ces derniers dans la circulation sanguine,
normalement présentes dans le plasma sanguin varient en taille et en
20 composition mais apparaissent toutes sous forme de microémulsion.

Les particules des lipoprotéines sont sphériques et contiennent un
noyau central de lipides non polaires (en majorité des triglycérides et des
esters de cholestérol) et une monocouche de surface de lipides polaires
(du cholestérol et majoritairement des phospholipides) et des protéines
25 dénommées apolipoprotéines (apo).

La plupart des composants protéiques de la monocouche de
surface, tous comme les phospholipides, ont des propriétés
amphipathiques. L'association des lipides polaires et des protéines avec
les lipoprotéines est ainsi réalisée via des forces hydrophobes grâce
30 auxquelles les chaînes d'acides gras et les chaînes latérales
d'acides aminés non polaires sont exclues de l'environnement aqueux.

La plupart des apolipoprotéines possèdent des régions amphipatiques en hélice (apolipoprotéine A-I, A-II, A-IV, C-I, C-II, C-III et E).

La densité des particules de lipoprotéine est inversement proportionnelle à leur taille, leur densité reflétant les quantités relatives de lipides non polaires de faible densité contenues dans le noyau et de protéines de surface de haute densité présentes.

Parmi les lipoprotéines de grande taille on connaît les chylomicrons, sécrétés par les entérocytes, dans lesquels l'apo B-48 est majoritaire, et les VLDL, sécrétées par les hépatocytes, qui contiennent la protéine apo B-100.

Les classes de lipoprotéines les plus petites, LDL et HDL, contiennent majoritairement des esters de cholestérol dans leur noyau. Les formes matures de ces particules ne sont pas sécrétées directement à partir des cellules mais sont plus particulièrement produites par des voies métaboliques au sein du plasma sanguin.

Les particules LDL représentent les produits finaux du métabolisme des particules VLDL.

Certains composants des particules HDL sont dérivés à partir des chylomicrons.

Ainsi, les lipoprotéines de haute densité (HDL) sont l'une des quatre classes majeures de lipoprotéines qui circulent dans le plasma sanguin.

Ces lipoprotéines sont impliquées dans différentes voies métaboliques telles que le transport lipidique, la formation des acides biliaires, la stéroïdogénèse, la prolifération cellulaire et en outre interfèrent avec les systèmes de protéinase plasmatique.

Les HDL sont de parfaits accepteurs de cholestérol libre et, en combinaison avec les protéines de transfert d'ester de cholestérol (CETP), la lipoprotéine lipase (LPL), la lipase hépatique (HL) et la lécithine : cholestérol acyltransférase (LCAT), jouent un rôle majeur dans le transport inverse du cholestérol, c'est à dire le transport du cholestérol en excès dans les cellules périphériques vers le foie pour son élimination de l'organisme sous forme d'acide biliaire.

Les HDL ne transportent pas seulement le cholestérol des cellules périphériques vers le foie mais le distribuent également aux cellules productrices de stéroïdes ou aux cellules périphériques appauvries en cholestérol.

5 Les précurseurs des HDL sont sécrétés sous forme discoidale à partir de l'intestin et du foie, à partir desquels les particules sphériques sont formées grâce à la formation d'esters de cholestérol qui migrent au cœur de la particule lipoprotéique.

10 Les particules naissantes de HDL contiennent des apo A-I et apo A-IV, alors que les particules de HDL hépatiques naissantes sont riches en apo A-I, apo E et apo A-II.

La partie lipidique de ces particules consiste en des phospholipides et de petites quantités de cholestérol libre et de triglycéride.

15 Il a été démontré que les HDL jouent un rôle central dans le transport du cholestérol des tissus périphériques vers le foie.

L'excès de cholestérol non estérifié des cellules périphériques riches en cholestérol est capté par les HDL et subit une estérification par l'action de la LCAT. Ces HDL enrichies en esters de cholestérol sont
20 captées par des protéines de liaison ou récepteurs des HDL à la surface des hépatocytes et y délivrent leurs esters de cholestérol.

Le rôle protecteur des HDL dans le transport inverse du cholestérol est confirmée par les études épidémiologiques démontrant une relation inverse entre les concentrations de cholestérol dans ces
25 HDL et le risque d'apparition de maladies coronariennes, ou encore par des observations selon lesquelles les HDL acceptent efficacement les excès de cholestérol intracellulaire à partir de types cellulaires variés.

Les lipoprotéines athérogènes sont ingérées par les macrophages ou cellules périphériques et dégradées dans les lysosomes. Le
30 cholestérol est relargué des lysosomes et est ré-estérifié dans le compartiment cytoplasmique.

Il a été en particulier montré que les HDL riches en apo A-I stimulent les flux de cholestérol à partir des macrophages ou des cellules périphériques vers le compartiment extracellulaire, suite à une

interaction avec des protéines de liaison des HDL présentes à la surface de ces cellules.

5 Diverses maladies liées à une déficience en HDL ont été décrites; comprenant la maladie de Tangier, la déficience en HDL et la déficience en LCAT.

10 La déficience impliquée dans la maladie de Tangier est reliée à un déficit cellulaire dans la translocation du cholestérol cellulaire dans lequel les précurseurs de HDL sont dégradés dans les lysosomes. Néanmoins, pour la maladie de Tangier, la nature exacte du déficit n'a pas encore été précisément définie.

15 Dans la maladie de Tangier, ce déficit cellulaire conduit à une perturbation du métabolisme lipoprotéique. Les HDL n'incorporant pas de cholestérol à partir des cellules périphériques et ne pouvant pas être métabolisée correctement, sont éliminées rapidement de l'organisme. La concentration plasmatique en HDL de ces patients est donc
20 extrêmement réduite et les HDL n'assurent plus le retour du cholestérol vers le foie. Ce cholestérol s'accumule dans ces cellules périphériques et provoquent des manifestations cliniques caractéristiques telles que la formation d'amygdales orangées. De plus, d'autres perturbations lipoprotéiques comme une surproduction de triglycérides ainsi qu'une synthèse et un catabolisme intracellulaire accrus des phospholipides sont observées.

25 La maladie de Tangier, dont les symptômes ont été décrits ci-dessus, est classée parmi les affections familiales liées au métabolisme des HDL qui sont les plus couramment détectées chez les patients affectés de maladies coronariennes.

30 De nombreuses études ont montré qu'un niveau réduit de cholestérol HDL est un excellent facteur de risque permettant de dépister une affection coronarienne.

Dans ce contexte, des syndromes liés aux déficiences en HDL ont présenté un intérêt accru durant la décennie passée du fait qu'elles permettent d'accroître la compréhension du rôle des HDL dans l'athérogénèse.

Plusieurs mutations dans le gène apo A-I ont été caractérisées. Ces mutations sont rares et peuvent conduire à une absence de production de apo A-I.

Des mutations dans les gènes codant pour la lipoprotéine lipase (LPL) ou son activateur apoC-II sont associées avec des hypertriglycéridémies sévères et des niveaux de HDL-c fortement réduits.

Des mutations dans le gène codant pour l'enzyme lécithine: cholestérol, acyltransférase (LCAT) sont également associées à une déficience sévère en HDL.

Il existe donc un besoin croissant dans l'état de la technique d'identifier des gènes impliqués dans le métabolisme du cholestérol et/ou des lipoprotéines, et en particulier de gènes associés à des dysfonctionnements du transport inverse du cholestérol des cellules périphériques vers le foie.

Récemment, une étude de la ségrégation de différentes formes alléliques de 343 marqueurs microsatellites répartis sur l'ensemble du génome et distants entre eux en moyenne de 10,3 cM a été réalisée.

L'étude de liaison (linkage) a porté sur une famille bien caractérisée sur onze générations, dont de nombreux membres sont affectés par la maladie de Tangier, la famille comportant cinq lignées de consanguinité.

Cette étude a permis d'identifier une région localisée dans le locus 9q31 du chromosome 9 humain statistiquement associé à l'affection (Rust S. et al., Nature Genetics, vol. 20, Septembre 1998, pages 96-98).

Toutefois, l'étude de RUST et al. définit seulement une large région du génome dont des altérations sont susceptibles d'être associées à la maladie de Tangier. Il est simplement précisé que la région du locus 9q31-34 concernée contient des ESTs mais aucun gène connu.

Il a désormais été montré qu'une région d'environ 15cM située dans le locus 9q31-34 chez l'homme était associée, de manière générale, à des déficiences familiales en HDL.

Plus particulièrement, on a montré selon l'invention que des molécules d'ARN messenger étaient exprimées à partir de séquences localisées dans le génome dans une région d'environ 15 cM centrée sur le marqueur microsatellite D9S1784, identifié comme le marqueur
5 microsatellite donnant la liaison génétique la plus forte avec la déficience familiale en HDL et la maladie de Tangier.

De plus, la région 9q31-34 considérée contient potentiellement des gènes susceptibles de jouer un rôle dans l'initiation ou le développement de diverses maladies, telles que :

- 10 - les maladies osseuses comme les chondrosarcomes myxoides, des retards mentaux liés aux anomalies du chromosome 9 (MRD) ;
- les maladies rénales comme les néphrophtisies infantiles (NPH2) ;
- les maladies musculaires, comme la dystrophie musculaire des ceintures (LGMD2H) ;
- 15 - les maladies psychiatriques comme la schizophrénie ;
- les maladies digestives comme la maladie de Hirschprung liée à RET (SHSCR2).

Du fait de la localisation des gènes candidats positionnels dans cette intervalle chromosomique, les ARN messagers et les polypeptides
20 correspondants, tels qu'isolés et caractérisés selon l'invention, sont potentiellement impliqués dans certaines des pathologies humaines décrites ci-dessus, soit encore dans d'autres pathologies également liées génétiquement à cette région du chromosome 9.

Pour certaines des séquences comprises dans ces ARN
25 messagers isolées et caractérisées par le demandeur, une phase de lecture ouverte putative a été déterminée et la séquence de la protéine correspondante en a été déduite. Les polypeptides correspondants sont potentiellement impliqués dans une affection liée au métabolisme des lipoprotéines, plus particulièrement liée à un déficit du transport inverse
30 du cholestérol.

DESCRIPTION DETAILLEE DE L'INVENTION

La présente invention décrit ainsi des polynucléotides et des polypeptides dont une altération de la séquence ou de l'expression est
5 potentiellement associée à un déficit dans le métabolisme des lipoprotéines plasmatiques, plus particulièrement à un déficit dans le transport inverse des HDL .

La présente invention décrit aussi des polynucléotides et des polypeptides dont une altération de la séquence ou de l'expression est
10 potentiellement associée à des maladies génétiquement liées au locus 9q31-34 du chromosome 9.

DEFINITIONS GENERALES

15 Le terme " isolé " au sens de la présente invention désigne un matériel biologique (acide nucléique ou protéine) qui a été soustrait à son environnement originel (l'environnement dans lequel il est localisé naturellement).

20 Par exemple un polynucléotide présent à l'état naturel dans une plante ou un animal n'est pas isolé. Le même polynucléotide séparé des acides nucléiques adjacents au sein desquels il est naturellement inséré dans le génome de la plante ou l'animal est considéré comme " isolé " .

Un tel polynucléotide peut être inclus dans un vecteur et/ou un tel
25 polynucléotide peut être inclus dans une composition et demeurer néanmoins à l'état isolé du fait que le vecteur ou la composition ne constitue pas son environnement naturel.

Le terme " purifié " ne nécessite pas que le matériel soit présent sous une forme de pureté absolue, exclusive de la présence d'autres
30 composés. Il s'agit plutôt d'une définition relative.

Un polynucléotide est à l'état " purifié " après purification du matériel de départ ou du matériel naturel d'au moins un ordre de
35 grandeur, de préférence 2 ou 3 et préférentiellement 4 ou 5 ordres de grandeur.

Aux fins de la présente description, l'expression " séquence nucléotidique " peut être employée pour désigner indifféremment un

polynucléotide ou un acide nucléique. L'expression "séquence nucléotidique" englobe le matériel de génétique lui-même et n'est donc pas restreinte à l'information concernant sa séquence.

Les termes "acide nucléique", "polynucléotide";
5 "oligonucléotide" ou encore "séquence nucléotidique" englobent des séquences d'ARN, d'ADN, d'ADNc ou encore des séquences hybrides ARN/ADN de plus d'un nucléotide, indifféremment sous la forme simple chaîne ou sous la forme de duplex.

Le terme "nucléotide" désigne à la fois les nucléotides naturels
10 (A, T, G, C) ainsi que des nucléotides modifiés qui comprennent au moins une modification telle que (1) un analogue d'une purine, (2) un analogue d'une pyrimidine, ou (3) un sucre analogue, des exemples de tels nucléotides modifiés étant décrits par exemple dans la demande PCT N°WO 95/04 064.

15 Aux fins de la présente invention, un premier polynucléotide est considéré comme étant "complémentaire" d'un second polynucléotide lorsque chaque base du premier nucléotide est appariée à la base complémentaire du second polynucléotide dont l'orientation est inversée. Les bases complémentaires sont A et T (ou a et U), ou C et G.

20 Par "variant" d'un acide nucléique selon l'invention, on entendra un acide nucléique qui diffère d'une ou plusieurs bases par rapport au polynucléotide de référence. Un acide nucléique variant peut être d'origine naturel, tel qu'un variant allélique retrouvé naturellement, ou peut être aussi un variant non naturel obtenu par exemple par des
25 techniques de mutagénèse.

En général, les différences entre l'acide nucléique de référence et l'acide nucléique variant sont réduites de telle sorte que les séquences nucléotidiques de l'acide nucléique de référence et de l'acide nucléique variant sont très proches et, dans de nombreuses régions, identiques.
30 Les modifications de nucléotides présentent dans un acide nucléique variant peuvent être silencieuses, ce qui signifie qu'elles n'altèrent pas les séquences d'acides aminés codées par ledit acide nucléique variant.

Cependant, les changements de nucléotides dans un acide nucléique variant peuvent aussi résulter dans des substitutions,
35 additions, délétions dans le polypeptide codé par l'acide nucléique

variant par rapport aux peptides codés par l'acide nucléique de référence. En outre, des modifications de nucléotides dans les régions codantes peuvent produire des substitutions, conservatives ou non conservatives dans la séquence d'acides aminés.

5 De préférence, les acides nucléiques variants selon l'invention codent pour des polypeptides qui conservent sensiblement la même fonction ou activité biologique que le polypeptide de l'acide nucléique de référence ou encore la capacité à être reconnus par des anticorps dirigés contre les polypeptides codés par l'acide nucléique initial.

10 Certains acides nucléiques variants coderont ainsi pour des formes mutées des polypeptides dont l'étude systématique permettra de déduire des relations structure activité des protéines en question. La connaissance de ces mutations par rapport à la maladie étudiée est fondamentale puisqu'elle permet de comprendre la cause moléculaire de la pathologie.

15 On entendra par "fragment" un acide nucléique de référence selon l'invention, une séquence nucléotidique de longueur réduite par rapport à l'acide nucléique de référence et comprenant, sur la partie commune, une séquence en nucléotides identique à l'acide nucléique de référence.

20 Un tel "fragment" d'acide nucléique selon l'invention peut être le cas échéant, compris dans un polynucléotide plus grand duquel il est constitutif.

25 De tels fragments comprennent, ou alternativement consistent en, des oligonucléotides de longueur allant de 8, 10, 12, 15, 18, 20 à 25, 30, 40, 50, 70, 80, 100, 200, 500, 1000 ou 1500 nucléotides consécutifs d'un acide nucléique selon l'invention.

30 Par "variant" d'un polypeptide selon l'invention, on entendra principalement un polypeptide dont la séquence d'acides aminés contient une ou plusieurs substitutions, additions ou délétions d'au moins un résidu d'acide aminé, par rapport à la séquence d'acides aminés du polypeptide de référence, étant entendu que les substitutions d'acides aminés peuvent être indifféremment conservatives ou non conservatives.

Par " fragment " d'un polypeptide selon l'invention, on entendra un polypeptide dont la séquence d'acides aminés est plus courte que celle du polypeptide de référence et qui comprend sur toute la partie commune avec ces polypeptides de référence, une séquence en acides aminés identique.

De tels fragments peuvent, le cas échéant, être compris au sein d'un polypeptide plus grand duquel ils font partie.

De tels fragments d'un polypeptide selon l'invention peuvent avoir une longueur de 10, 15, 20, 30 à 40, 50, 100, 200 ou 300 acides aminés.

Le " pourcentage d'identité " entre deux séquences de nucléotides ou d'acides aminés, au sens de la présente invention, peut être déterminé en comparant deux séquences alignées de manière optimale, à travers une fenêtre de comparaison.

La partie de la séquence nucléotidique ou polypeptide dans la fenêtre de comparaison peut ainsi comprendre des additions ou des délétions (par exemple des " gaps ") par rapport à la séquence de référence (qui ne comprend pas ces additions ou ces délétions) de manière à obtenir un alignement optimal des deux séquences.

Le pourcentage est calculé en déterminant le nombre de positions auquel une base nucléique ou un résidu d'acide aminé identique est observé pour les deux séquences (nucléique ou peptidique) comparées, puis en divisant le nombre de positions auquel il y a identité entre les deux bases ou résidus d'acides aminés par le nombre total de positions dans la fenêtre de comparaison, puis en multipliant le résultat par 100 afin d'obtenir le pourcentage d'identité de séquence.

L'alignement optimal des séquences pour la comparaison peut être réalisé de manière informatique à l'aide d'algorithmes connus contenus dans le package de la Société WISCONSIN GENETICS SOFTWARE PACKAGE, GENETICS COMPUTER GROUP (GCG), 575 Science Doctor, Madison, WISCONSIN.

A titre d'illustration, le pourcentage d'identité de séquence pourra être effectué à l'aide du logiciel BLAST (versions BLAST 1.4.9 de mars 1996, BLAST 2.0.4 de février 1998 et BLAST 2.0.6 de septembre 1998), en utilisant exclusivement les paramètres par défaut (S. F Altschul et al, J. Mol. Biol. 1990 215 : 403-410, S. F Altschul et al, Nucleic Acids Res.

1997 25 : 3389-3402). Blast recherche des séquences
similaires/homologues à une séquence " requête " de référence, à l'aide
de l'algorithme d'Altschul et al. La séquence requête et les bases de
données utilisées peuvent être peptidiques ou nucléiques, toute
5 combinaison étant possible.

Par " conditions d'hybridation de forte stringence " au sens de la
présente invention, on entendra les conditions suivantes :

10 1- Compétition des membranes et PRE HYBRIDATION :

- Mélanger : 40µl ADN sperme de saumon (10mg/ml)
+ 40 µl ADN placentaire humain (10mg/ml)
- 15 - Dénaturer 5 mn à 96°C, puis plonger dans la glace le mélange.
- Oter le SSC 2X et verser 4 ml de mix formamide dans le tube
d'hybridation contenant les membranes.
- 20 - Ajouter le mélange des deux ADNs dénaturés.
- Incubation à 42°C pendant 5 à 6 heures, avec rotation.

2- Compétition de la sonde marquée :

- 25
- Ajouter à la sonde marquée et purifiée 10 à 50 µl ADN Cot I, selon la
quantité de repeats.
 - Dénaturer 7 à 10 mn à 95°C.
 - 30 - Incuber à 65°C pendant 2 à 5 heures.

3- HYBRIDATION :

- Oter le mix de pré hybridation.

5

- Mélanger 40 µl ADN sperme de saumon + 40 µl ADN placentaire humain ; dénaturer 5 mn à 96°C, puis plonger dans la glace.

10 - Ajouter dans le tube d'hybridation 4 ml de mix formamide, le mélange des deux ADN et la sonde marquée/ADN Cot I dénaturée.

- Incuber 15 à 20 heures à 42°C, avec rotation.

4- Lavages :

15

- Un lavage à température ambiante dans du SSC 2X, pour rincer.

- 2 fois 5 minutes à température ambiante SSC 2X et SDS 0,1% à 65°C.

- 2 fois 15 minutes à 65°C SSC 1X et SDS 0,1% à 65°C.

20 Envelopper les membranes dans du Saran et exposer.

25 Les conditions d'hybridation décrites plus haut sont adaptées à l'hybridation dans des conditions de forte stringence, d'une molécule d'acide nucléique d'une longueur variable de 20 nucléotides à plusieurs centaines de nucléotides.

30 Il va sans dire que les conditions d'hybridation ci-dessus décrites peuvent être adaptées en fonction de la longueur de l'acide nucléique dont l'hybridation est recherchée ou du type de marquage choisi, selon des techniques connues de l'homme du métier.

Les conditions convenables d'hybridation peuvent par exemple être adaptées selon l'enseignement contenu dans l'ouvrage de HAMES et HIGGINS (1985) ou encore dans l'ouvrage de F.AUSUBEL et al (1999).

5

DETAIL DES ACIDES NUCLEIQUES ET DES POLYPEPTIDES SELON L'INVENTION

Une brève description des séquences d'acides nucléiques et des
10 séquences d'acides aminés selon l'invention est représentée dans le Tableau I, à la suite des exemples.

Gène GS9002S31.

Acide(s) nucléique(s)

15 Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS9002S31.

La séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°1.

La séquence SEQ ID N°1 a une longueur de 552 nucléotides.

20 Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

Les analyses d'expression du transcrit de séquence SEQ ID N°1 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis
25 de montrer que le gène GS9002S31 était exprimé dans le cerveau fœtal, le foie et le placenta.

Gène GS910331.***Acide(s) nucléique(s)***

Il a été isolé selon l'invention deux séquences d'ARN messenger correspondant à un transcrit du gène désigné ici sous le n° GS910331.

5 La première séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°2.

La séquence SEQ ID N°2 a une longueur de 1246 nucléotides.

La seconde séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°3.

10 La séquence SEQ ID N° 3 a une longueur de 3035 nucléotides.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110 et Version 115).

15 Les analyses d'expression du transcrit de séquence SEQ ID N°2 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS310331 était exprimé dans le cerveau foetal.

De plus, une analyse de l'expression du transcrit par Northern blot, à l'aide respectivement des sondes de séquences SEQ ID N°82 et
20 83, a révélé la présence de transcrits dans le blot commercialisé par la Société Clontech (Ref. N° 7759-1).

La taille des transcrits détectés avec la sonde de séquence SEQ ID N° 82 est respectivement de 1,65 kb dans le foie et le cœur et de 1,4 kb dans le cerveau.

25 La taille des transcrits détectés avec la sonde de séquence SEQ ID N° 83 est respectivement de 1,65 kb et 2,4 kb dans le cœur et de 1,65 kb dans le foie.

30 Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus

particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

5 **Gène GS914554**

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS94554.

10 La séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°4.

La séquence SEQ ID N°4 a une longueur de 1479 nucléotides.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

15 Les analyses d'expression du transcrit de séquence SEQ ID N°4 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS94554 était exprimé dans le cerveau fœtal, le placenta et le foie.

20 De plus, une analyse de l'expression du transcrit par Northern blot, selon le protocole décrit de l'Exemple 1, à l'aide respectivement de la sonde de séquence SEQ ID N°58, a révélé la présence de transcrits dans le blot commercialisé par la Société Clontech (Ref. N° 7759-1).

25 La taille des transcrits détectés avec la sonde de séquence SEQ ID N°84 est respectivement de :

- 1,0, 1,3, 1,7 et 2,8 kb dans le pancréas et le placenta ;
- 1,0, 1,3 et 1,7 kb dans le rein, le muscle squelettique, le coeur et le foie ;
- 1,7 kb dans le cerveau et le poumon.

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Gène GS914739.

Acide(s) nucléique(s)

Il a été isolé selon l'invention deux séquences d'ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS14739.

La première séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N 5.

La séquence SEQ ID N°5 a une longueur de 5169 nucléotides.

Aucune homologie de séquence avec la séquence SEQ ID N° 5 n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

La seconde séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N 6.

La séquence SEQ ID N°6 a une longueur de 7723 nucléotides. Cette séquence comprend un cadre ouvert de lecture (ORF) allant du nucléotide en position 121 au nucléotide en position 1517 de la séquence SEQ ID N° 6. Le codon d'initiation de la traduction débute au nucléotide en position 132 de la séquence SEQ ID N° 6. La séquence codante débute au nucléotide en position 132 et se termine au nucléotide en position 1517 de la séquence SEQ ID N° 6. La séquence SEQ ID N°6 comprend un signal de polyadénylation de séquence " ATTAAA " débutant au nucléotide en position 7686 de la séquence SEQ ID N°6. Le motif de Kozak de séquence " CCA CTC GCC ATG " débute au nucléotide en position 123 de la séquence SEQ ID N°6.

Deux homologies de séquence de 100% ont été retrouvées lors d'une recherche dans la base de données GenBank (Version 115, N° d'accès AF088031), respectivement :

- du nucléotide en position 1 au nucléotide en position 146 de la séquence SEQ ID N°6 ; et
- du nucléotide en position 243 au nucléotide en position 573 de la séquence SEQ ID N° 6

Les analyses d'expression du transcrit de séquence SEQ ID N°5 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+de différents tissus ont permis de montrer que le gène GS914739 était exprimé dans le cerveau foetal.

De plus, une analyse de l'expression du transcrit par Northern blot, à l'aide de la sonde de séquence SEQ ID N°85, a révélé la présence de transcrits dans le blot commercialisé par la Société Clontech (Ref. N° 7759-1).

La taille des transcrits détectés avec la sonde de séquence SEQ ID N°85 est de 1 kb dans le cœur, le foie, le muscle squelettique et le rein.

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Polypeptide codé par l'acide nucléique de séquence SEQ ID N° 6

Le cadre ouvert de lecture de la séquence nucléotidique SEQ ID N° 6 code potentiellement pour un polypeptide de 461 acides aminés de longueur constituant la séquence SEQ ID N° 129.

Sur les régions nucléotidiques 240-1481 et 1511-1675 de la séquence SEQ ID N° 129, une identité de séquence d'environ 30% a été retrouvée avec dans les bases Genpept115, Swissprot38, trEMBL et PIR vace les
5 numéro d'accès suivants : AF035360 (homo), AF186461 (rattus), AF186460 (muspretus); AF196481 (homo sapiens), AF196480(musmusc) et avec des T09482(homme) et T09013(mouse) (ring finger Fxy). Quelques homologies de séquence ont été retrouvées également avec les numéros
10 d'accès suivants : DA191P20.2, A49656 et I49642.

Le polypeptide de séquence SEQ ID N° 129 est susceptible d'intervenir dans la régulation du flux du cholestérol, et plus particulièrement de la maladie de Tangier, dans des déficiences familiales en HDL, ou encore dans une maladie génétiquement liée au
15 locus 9q31-34 du chromosome 9.

Gène S915574.

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messenger correspondant à
20 un transcrit du gène désigné ici sous le n° GS915574.

La séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°7.

La séquence SEQ ID N°7 a une longueur de 1046 nucléotides.

Aucune identité de séquence n'a été retrouvée lors d'une
25 recherche dans la base de données GenBank (Version 110).

Les analyses d'expression du transcrit de séquence SEQ ID N°7 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+de différents tissus ont permis de montrer que le gène GS915574 était exprimé dans le cerveau foetal,

l'utérus, le cerveau, le cœur, la prostate, le foie fœtal, le foie, le placenta, le testicule et le rein.

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

10

Gène GS930321

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS930321.

15

La séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°8.

La séquence SEQ ID N°8 a une longueur de 280 nucléotides.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

20

Les analyses d'expression du transcrit de séquence SEQ ID N°8 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS930321 était exprimé dans le cerveau fœtal, le foie et le cœur.

25

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus

30

9q31-34 du chromosome 9.

Gène S931311.***Acide(s) nucléique(s)***

5 Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS931311.

La séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°9.

10 La séquence SEQ ID N°9 a une longueur de 479 nucléotides. Cette séquence comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 3 au nucléotide en position 98 de la séquence SEQ ID NO 9

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

15 Les analyses d'expression du transcrit de séquence SEQ ID N°9 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS931311 était exprimé dans le cerveau fœtal, le foie, le cœur, le placenta, le testicule et le rein.

20 Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus
25 9q31-34 du chromosome 9.

Polypeptide codé par l'acide nucléique de séquence SEQ ID NO 9

Le cadre ouvert de lecture partiel de la séquence d'acides nucléique SEQ ID N°9 code potentiellement pour un polypeptide de 32 aminoacides de longueur constituant la séquence SEQ ID N°130.

Aucune homologie significative avec BLAST n'a été retrouvée
5 avec les séquences référencées dans les bases de données Swissprot (version 36, dernière remise à jour du 3 mai 1999) et PRODOM: (domaines homologues obtenus dans Swissprot, version 34.2, novembre 1997).

Le polypeptide de séquence SEQ ID N° 130 est susceptible
10 d'intervenir dans la régulation du flux du cholestérol, et plus particulièrement de la maladie de Tangier, dans des déficiences familiales en HDL, ou encore dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

15 **Gène S934660.**

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS934660.

La séquence nucléique de l'ADNc correspondant à ce transcrit
20 constitue la séquence SEQ ID N°10.

La séquence SEQ ID N°10 a une longueur de 2599 nucléotides.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

Les analyses d'expression du transcrit de séquence SEQ ID N°10
25 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS934660 était exprimé dans le cerveau foetal.

De plus, une analyse de l'expression du transcrit par Northern blot, à l'aide de la sonde de séquence SEQ ID N°86, selon le protocole

décrit de l'Exemple 1, a révélé la présence de transcrits dans le blot commercialisé par la Société Clontech (Ref. N° 7759-1).

La taille des transcrits détectés avec la sonde de séquence SEQ ID N°86 est respectivement de :

- 5 - 1 kb, 2 kb, 3 kb et 7,5 kb dans le placenta ;
- 2, 3 et 7,5 kb dans le cœur ;
- 7,5 kb dans le rein, le pancréas, le muscle squelettique, le poumon et le cerveau.

- 10 Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

15

Gène GS938315.

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS938315.

- 20 La séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°11.

La séquence SEQ ID N°11 a une longueur de 222 nucléotides.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

- 25 Les analyses d'expression du transcrit de séquence SEQ ID N°11 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS938315 était exprimé dans le cerveau fœtal, le foie, le cœur et le rein.

30

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Gène GS93953

Acide(s) nucléique(s)

Il a été isolé selon l'invention deux ARNs messagers correspondant à un transcrit du gène désigné ici sous le n° GS93953.

La première séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°12.

La séquence SEQ ID N°12 a une longueur de 3422 nucléotides.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

La seconde séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N° 13.

La séquence SEQ ID N° 13 a une longueur de 5791 nucléotides. Elle comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 3 au nucléotide en position 554 de la séquence SEQ ID N° 13.

Quelques homologues de séquence ont été retrouvés entre la séquence SEQ ID N° 13 et les numéros d'accès suivants de la base de données GenBank (Version 116): AC013740.2, AC013783.2 et AF086175.1.

Les analyses d'expression du transcrit de séquence SEQ ID N°12 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces

analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS93953 était exprimé dans le cerveau foetal.

De plus, une analyse de l'expression du transcrit par Northern blot, , selon le protocole décrit de l'Exemple 1, à l'aide de la sonde de
5 séquence SEQ ID N°87, a révélé la présence de transcrits dans le blot commercialisé par la Société Clontech (Ref. N° 7759-1).

La taille des transcrits détectés avec la sonde de séquence SEQ ID N°87 est de 8 kb dans le cœur, le cerveau, le placenta, le poumon, le foie, le muscle squelettique, le rein et le pancréas.

10

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus
15 9q31-34 du chromosome 9.

Polypeptide codé par l'acide nucléique de séquence SEQ ID N° 13.

Le cadre de lecture ouvert partiel de la séquence d'acide
20 nucléique SEQ ID N°13 code potentiellement pour un polypeptide de 183 acides aminés de longueur constituant la séquence SEQ ID N° 131.

Il a été observé une homologie de 45% entre la région 6-162 de la séquence SEQ ID N° 131 et la séquence n° g3878571 (Z46381) de la base de données Gen Bank (Version 115) et la séquence n°
25 EM:Q21453 MO1F1.4 PROTEIN de la base de données trEMBL (Version de Août 1999).

Le polypeptide de séquence SEQ ID N° 131 est susceptible d'intervenir dans la régulation du flux du cholestérol, et plus particulièrement de la maladie de Tangier, dans des déficiences

familiales en HDL, ou encore dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Gène GS939874.

5 ***Acide(s) nucléique(s)***

Il a été isolé selon l'invention deux ARNs messagers correspondant à un transcrit du gène désigné ici sous le n° GS939874.

La première séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°14.

10 La séquence SEQ ID N°14 a une longueur de 2615 nucléotides.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

La deuxième séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N° 15.

15 La séquence SEQ ID N° 15 a une longueur de 2551 nucléotides. Elle comprend un cadre ouvert de lecture allant du nucléotide en position 50 au nucléotide en position 958 et une séquence codante allant du nucléotide en position 67 au nucléotide en position 958.

Une homologie de 99% d'identité sur 2044 nucléotides de la
20 séquence SEQ ID N° 15 a été observée avec la séquence de la base de données GenBank (Version 116) ayant le numéro d'accès AK001355.

Les analyses d'expression du transcrit de séquence SEQ ID N°14 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA⁺ de différents tissus ont permis
25 de montrer que le gène GS939874 était exprimé dans le cerveau fœtal, l'utérus, le cerveau, le cœur, la prostate, le foie fœtal, le foie, le placenta, le testicule et le rein.

Ce gène constitue un candidat positionnel causal d'une maladie
30 due à un dysfonctionnement du flux inverse du cholestérol, et plus

particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

5 ***Polypeptide codé par l'acide nucléique de séquence SEQ ID N° 15.***

Le cadre ouvert de lecture de l'acide nucléique de séquence SEQ ID N° 15 code potentiellement pour un polypeptide de 291 acides aminés de longueur constituant la séquence SEQ ID N° 132.

10 Une homologie de 35% sur 233 acides aminés (14-246 de SEQ ID N° 132) a été retrouvée avec la séquence référencée dans la base de données GenPept (Version 115) sous le n° g5832945 (AL117195).

Une homologie de 32% sur 245 acides aminés (30-274 de SEQ ID N° 132) a été retrouvée avec la séquence référencée dans la base de
15 données GenPept (Version 115) sous le n° g5832942 (AL117195).

Le polypeptide de séquence SEQ ID N° 132 est susceptible d'intervenir dans la régulation du flux du cholestérol, et plus particulièrement de la maladie de Tangier, dans des déficiences
20 familiales en HDL, ou encore dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Gène GS911370

Acide(s) nucléique(s)

25

Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS911370.

La séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°16.

30 La séquence SEQ ID N°16 a une longueur de 775 nucléotides.

Cette séquence comprend un cadre ouvert de lecture partiel (ORF) allant du nucléotide en position 1 au nucléotide en position 144 de la séquence SEQ ID N°16.

Une homologie en nucléotides a été retrouvée avec les
5 séquences suivantes :

- 96% d'homologie sur 229 pb (position 52-280 pb) avec la
séquence GenBank: gi|1022224| - Fragment Mse1
10 d'ADN génomique d'Homo sapiens, contenant un ilot
CpG (clone 92e10, lecture inverse de cpg92e10.rt1a)
- 100% d'homologie sur 145 pb (position 1-144 pb) avec la
séquence GenBank: gi|459833| de l'ARNm humain de la
15 sous unité bêta du complexe Sec61.

Les analyses d'expression du transcrit de séquence SEQ ID N°16
ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces
analyses effectuées à partir d'ARN polyA⁺ de différents tissus ont permis
20 de montrer que le gène GS911370 était exprimé dans le cerveau foetal.

De plus, une analyse de l'expression du transcrit par Northern
blot, selon le protocole décrit de l'Exemple 1, à l'aide respectivement des
sondes de séquence SEQ ID N° 88 et 89, a révélé la présence de
transcrits dans le blot commercialisé par la Société Clontech (Ref. N°
25 7759-1).

La taille des transcrits détectés avec la sonde de séquence SEQ
ID N° 88 ainsi qu'avec la sonde de séquence SEQ ID N°89 est de 7,4
kb dans le pancréas.

30 Ce gène constitue un candidat positionnel causal d'une maladie
due à un dysfonctionnement du flux inversé du cholestérol, et plus

particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

5 ***Polypeptide codé par l'acide nucléique de séquence SEQ ID N° 16***

Le cadre ouvert de lecture partiel de la séquence d'acides nucléiques SEQ ID N°16 code potentiellement pour un polypeptide de 48 aminoacides de longueur constituant la séquence SEQ ID N°133.

10 Homologies observées au niveau de la séquence protéique :
Cette ORF potentielle (48 aa) possède une identité sur 33 aa (position 16-48 aa) avec la protéine de transport sous unité bêta du complexe Sec61 humaine dont les séquences sont sp|P38391|, gb|AAA19639.1|, PIR: (Séquences PIR non redondantes, version 57)|S|S42410 et 18652
15 p34.2 (1) respectivement dans les bases de données Swissprot (version 36, dernière remise à jour du 3 mai 1999))Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999), PIR: (SEQUENCES PIR NON REDONDANTES, VERSION 57) non-redondante et PRODOM: (domaines homologues détectés dans
20 Swissprot, version 34.2, novembre 1997).

Cette identité est retrouvée dans les différentes bases de données protéiques (PIR: (SEQUENCES PIR NON REDONDANTES, VERSION 57), PRODOM: (domaines homologues détectés dans Swissprot, version 34.2, novembre 1997)) et les traductions de Genbank
25 et EMBL (TrEMBL (SP-TrEMBL, version 7, novembre 1998), Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999)).

Fonction putative :

30 Le complexe de protéines sec61 est un composant central de la machinerie cellulaire de translocation des protéines naissantes dans le

réticulum endoplasmique. Le gène GS911370 pourrait donc être, de part son homologie avec la sous unité beta du complexe sec61, un gène codant pour un nouveau composant de cette machinerie. Par conséquent son rôle possible dans la translocation des protéines et par
5 conséquent du trafic des protéines impliquées dans le mécanisme d'efflux du cholestérol en fait un gène d'intérêt dans l'étude du déficit observée chez les patients Tangier/FHD.

Le polypeptide de séquence SEQ ID N°133 est ainsi susceptible d'intervenir dans la régulation du flux de cholestérol, et plus
10 particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL.

Le polypeptide de séquence SEQ ID N°133 est donc susceptible d'intervenir dans une étape importante impliquée dans le transport inverse du cholestérol par les HDL.

15 Le polypeptide de séquence SEQ ID N° 133 est également susceptible d'intervenir dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Gène GS913920.

20 ***Acide(s) nucléique(s)***

Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS913920.

Une première séquence a été isolée et caractérisée ; il s'agit de la séquence nucléique de l'ADNc constituant la séquence SEQ ID N°17.

25 La séquence SEQ ID N°17 a une longueur de 491 nucléotides.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

30 Les analyses d'expression du transcrit de séquence SEQ ID N°17 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces

analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS913920 était exprimé dans le foie et le coeur.

A partir de la séquence SEQ ID N° 17, deux amorces nucléotidiques respectivement de séquences SEQ ID N° 102 et SEQ ID
5 N° 103 ont été synthétisées. Ces amorces nucléotidiques ont permis d'amplifier un ADNc du gène GS913920 constituant la séquence SEQ ID N° 18.

Les réactions d'amplification sont réalisées dans les conditions suivantes, applicables à l'ensemble des gènes candidats selon
10 l'invention pour lesquelles des amorces spécifiques sont décrites et mises en œuvre pour isoler des séquences du transcrit d'intérêt :

Chaque réaction de PCR est effectuée avec 400 µM de chaque dNTP, 0.5 µM de chaque amorce, 2.5 mM de MgCl₂, 50 ng d'ADN ou environ 25 ng d'ADNc et 2 unités de *Thermus aquaticus* (Taq) DNA
15 polymérase (Ampli Taq Gold; Perkin Elmer) en présence de son tampon. Les réactions sont réalisées en micro-plaques 96 puits, dans des thermo-cycleurs 9700 (Perkin Elmer). Après une première dénaturation à 94°C pendant 10 min, un programme de 30 cycles est appliqué : dénaturation de 30 s. à 94°C, hybridation de 30 s. à 64°C (2 cycles),
20 61°C (2 cycles), 58°C (2 cycles) et 55°C (28 cycles), élongation de 1 min/kb à 72°C. Le programme s'achève par une élongation de 7 min à 72°C.

La séquence nucléique SEQ ID N° 18 a une longueur de 293
25 nucléotides. Elle comprend un cadre de lecture ouvert (ORF) partiel allant du nucléotide en position 227 au nucléotide en position 293.

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences

familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Polypeptide codé par l'acide nucléique de séquence SEQ ID NO 18

5

Le cadre ouvert de lecture de la séquence d'acides nucléiques SEQ ID N°18 code potentiellement pour un polypeptide de 22 aminoacides de longueur constituant la séquence SEQ ID N°134.

Aucune homologie significative avec BLAST n'a été retrouvée
10 avec les séquences référencées dans les bases de données Genpept: (traduction de Genbank v115), TrEMBL (SP-TrEMBL, Version Août 1999), Swissprot (Version 38) et PIR: (SEQUENCES PIR NON REDONDANTES, Version 62-Septembre 1999).

Le polypeptide de séquence SEQ ID N°134 est susceptible
15 d'intervenir dans la régulation du flux de cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL.

Le polypeptide de séquence SEQ ID N°134 est donc susceptible
20 d'intervenir dans une étape importante impliquée dans le transport inverse du cholestérol par les HDL.

Le polypeptide de séquence SEQ ID N° 134 est également susceptible d'intervenir dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

25 **Gène GS91437.**

Acide(s) nucléique(s)

Il a été isolé selon l'invention deux ARNs messagers correspondant à un transcrit du gène désigné ici sous le n° GS91437.

La première séquence nucléique de l'ADNc correspondant à ce
30 transcrit constitue la séquence SEQ ID N°19.

La séquence SEQ ID N°19 a une longueur de 2442 nucléotides.

Cette séquence comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 2 au nucléotide en position 286 de la séquence SEQ ID N°19.

5 Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

La seconde séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N° 20.

10 La séquence SEQ ID N°20 a une longueur de 4608 nucléotides. Elle comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 1 au nucléotide en position 327.

Il a été trouvé les homologues de séquence suivantes dans la base de données GenBank (Version 116) avec la séquence SEQ ID N° 20 :

- 15 • 100% d'homologie sur 213 pb (positions [85-297]) avec les positions [2807-2595] de la séquence g7020279 (AK000294) Homo sapiens cDNA FLJ20287 fis, clone HEP04390 Length = 3043; déposée le 22-02-2000 dans Genbank.
- 20 • 88% sur 219 pb (positions [82-300]) avec la séquence g3850048 (AJ 004828) Mus musculus clone XX-BAC394, Length = 170351*** SEQUENCING IN PROGRESS ***, in unordered pieces;

25 Les analyses d'expression du transcrit de séquence SEQ ID N°19 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA⁺ de différents tissus ont permis de montrer que le gène GS91437 était exprimé dans le cerveau fœtal, le foie, le cœur, la prostate, le placenta, l'utérus, le testicule, le rein, le muscle squelettique.

30 Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences

familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Polypeptide codé par l'acide nucléique de séquence SEQ ID N°19

5

Le cadre ouvert de lecture partiel de la séquence d'acides nucléiques SEQ ID N°19 code potentiellement pour un polypeptide de 95 aminoacides de longueur constituant la séquence SEQ ID N°135.

Aucune homologie significative avec BLAST n'a été retrouvée
10 avec les séquences référencées dans les bases de données Swissprot (version 36, dernière remise à jour du 3 mai 1999), PRODOM: (domaines homologues détectés dans Swissprot, versions 34.2 et 38,) Genpept: (traduction de Genbank v110 et 111 ainsi que 115,), PIR: (SEQUENCES PIR NON REDONDANTES, VERSION 57), PDB:
15 (PROTEIN DATA BANK, FEVRIER 1999) et TrEMBL (SP-TrEMBL, version 7, novembre 1998).

Le polypeptide de séquence SEQ ID N° 135 est susceptible d'intervenir dans la régulation du flux du cholestérol, et plus particulièrement de la maladie de Tangier, dans des déficiences
20 familiales en HDL, ou encore dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Polypeptide codé par l'acide nucléique de séquence SEQ ID N° 20

25

Le cadre ouvert de lecture (ORF) contenu dans l'acide nucléique de séquence SEQ ID N° 20 code potentiellement pour un polypeptide de 108 acides aminés constituant la séquence SEQ ID N° 136.

Aucune homologie de séquence n'a été retrouvé avec les bases de données Swissprot (Version 38), Genpept (Version 115), PIR
30 (Version 62, Septembre 1999) et trEMBL (Version de Août 1999).

Le polypeptide de séquence SEQ ID N° 136 est susceptible d'intervenir dans la régulation du flux du cholestérol, et plus particulièrement de la maladie de Tangier, dans des déficiences familiales en HDL, ou encore dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Gène GS91507.

Acide(s) nucléique(s)

Il a été isolé selon l'invention deux ARNs messagers correspondant à un transcrit du gène désigné ici sous le n° GS91507.

La première séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°21.

La séquence SEQ ID N°21 a une longueur de 1627 nucléotides.

Cette séquence comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 1 au nucléotide en position 640 de la séquence SEQ ID N°21.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

La deuxième séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N° 22.

La séquence SEQ ID N° 22 a une longueur de 2333 nucléotides. Elle comprend un cadre ouvert de lecture (ORF) complet allant du nucléotide en position 368 au nucléotide en position 1348. Le début du codon d'initiation de la traduction est localisé sur le nucléotide en position 371 de la séquence SEQ ID N° 22. La séquence codante débute au nucléotide en position 371 et se termine au nucléotide en position 1348.

La séquence SEQ ID N° 22 possède une homologie avec la séquence suivante référencée dans la base de données GenBank (Version 116) suivante :

Identité nucléique de 99% sur 2316 pb (position 115 à 2420 pb) avec Homo sapiens cDNA FLJ20300 fis, clone HEP06465 (2331 pb). Accession AK000307

5

Les analyses d'expression du transcrit de séquence SEQ ID N°21 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS91507 était exprimé dans le cerveau foetal.

10

De plus, une analyse de l'expression du transcrit par Northern blot, selon le protocole décrit de l'Exemple 1, à l'aide de la sonde de séquence SEQ ID N°90, a révélé la présence de transcrits dans le blot commercialisé par la Société Clontech (Ref. N° 7759-1).

La taille des transcrits détectés avec la sonde de séquence SEQ ID N°90 est respectivement de :

- 2 kb et 7,5 kb dans le pancréas, le rein, le muscle squelettique, le poumon, le placenta et le cerveau.

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL.

Polypeptide codé par l'acide nucléique de séquence SEQ ID NO 21

Le cadre ouvert de lecture de la séquence d'acides nucléiques SEQ ID N°21 code potentiellement pour un polypeptide de 213 aminoacides de longueur constituant la séquence SEQ ID N°137.

Homologies observées au niveau de la séquence protéique:
Cette ORF potentielle de 213 aa possède un domaine ayant des homologues de type motif avec différentes protéines telles que :

30

- 29% d'homologie sur 155 aa (position 4-158 aa) avec des séquences de Swissprot (version 36, dernière remise à jour du 3 mai 1999) sp|Q10022| et avec des séquences de PRODOM: (domaines homologues détectés dans Swissprot, version 34.2, novembre 1997) 28705 p34.2 (1) YSX3_CAEEL - domaine de la protéine hypothétique (39KD) T28D9.3 localisé sur le chromosome II.
- 25% d'homologie sur 127 aa (position 6-132 aa) avec des séquences de PIR: (SEQUENCES PIR NON REDONDANTES, VERSION 57) non-redondante: PIR: (Séquences PIR non redondantes, version 57)|S|S66668 - protéine induite par le peroxyde d'hydrogène - (fragment d'une séquence de souris).

15

De plus, des homologies avec des séquences codantes traduites de Genbank et EMBL (TrEMBL (SP-TrEMBL, version 7, novembre 1998), Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999)) montreraient des annotations avec des protéines

20 potentielles de type " acide phosphatidique phosphatase "

- 34% d'homologie sur 200 aa (position 6-205 aa) avec des séquences de SP-TrEMBL (SP-TrEMBL, version 7, novembre 1998): sp|P97544|P97544 - protéine transmembranaire du réticulum endoplasmique.
- 33% d'homologie sur 204 aa (position 6-209 aa) avec des séquences de Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999): gi|4105139|- acide phosphatidique phosphatase-beta,

30

type2 ; phosphatidate phosphohydrolase ; phospholipide phosphatase chez l'Homme. ; et avec des séquences de Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999): gi|3047173|
 5 homologue de l'acide phosphatidique phosphohydrolase de l'Homme, et avec gi|2467300| |- acide phosphatidique phosphatase 2b.

- 31% d'homologie sur 203 aa (position 6-208 aa) avec des séquences de Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999): gi|1487873||- acide phosphatidique phosphatase de souris

- 31% d'homologie sur 203 aa (position 6-208 aa) avec des séquences de SP-TrEMBL (SP-TrEMBL, version 7, novembre 1998): sp|Q61469|Q61469 |- acide phosphatidique phosphatase 2A

20 ***Polypeptide codé par l'acide nucléique de séquence SEQ ID N° 22.***

Le cadre ouvert de lecture de la séquence d'acide nucléique SEQ ID N°22 code potentiellement pour un polypeptide de 325 aminoacides de longueur constituant la séquence SEQ ID N°138.

- 25 Homologies observées au niveau de la séquence proteique
- Homologies avec des séquences codantes traduites de Genbank (Version 116) et EMBL (TrEMBL (Version de Août 1999), Genpept [Version 115]) montreraient des annotations avec des protéines potentielles de type " acide phosphatidique phosphatase ".

- 30% d'homologie sur 316 aa (position 2-317 aa) avec SP-trEMBL: sp|P97544|P97544 ER TRANSMEMBRANE PROTEIN
- 5 • 30% d'homologie sur 320 aa (position 2-321aa) avec Genpept116 gi|4105139|AF043329 type-2 phosphatidic acid phosphatase-beta; phosphatidate phosphohydrolase; phospholipid phosphatase [Homo sapiens] et avec Genpept116: gi|3047173|AF01786 phosphatidic acid phosphohydrolase homolog [Homo sapiens] et avec gi|2467300|AB000889 phosphatidic acid phosphatase 2b
- 10 • 30% d'homologie sur 316 aa (2-317 aa) avec Genpept116 gi 1684745 Y07783 transmembrane protein [Rattus norvegicus] Length = 312 aa
- 15 • 30% d'homologie sur 320 aa (position 2-317 aa) avec SP-trEMBL EM:O14495 PHOSPHATIDIC ACID PHOSPHATASE 2B. Length = 311 aa

20

Fonction putative des polypeptides de séquence SEQ ID N° 137 et 138:

Le gène GS91507 présente un intérêt pour l'étude fonctionnelle de la maladie de Tangier et du FHD en ce sens qu'il s'agit à la fois d'un
25 candidate positionnel et d'un gène codant pour une nouvelle protéine dont les fonctions putatives (phosphatidic acid phosphatase) pourraient jouer un rôle dans la cascade de signalisation intracellulaire liée à l'efflux du cholestérol intracellulaire médié par les particules HDL, impliqué dans la maladie de Tangier et la déficience familiale en HDL (FHD).

Les polypeptides de séquence SEQ ID N° 137 et 138 est susceptible d'intervenir dans la régulation du flux du cholestérol, et plus particulièrement de la maladie de Tangier, dans des déficiences familiales en HDL, ou encore dans une maladie génétiquement liée au
5 locus 9q31-34 du chromosome 9.

Gène GS915231.

Acide(s) nucléique(s)

Il a été isolé selon l'invention deux ARNs messagers
10 correspondant à un transcrit du gène désigné ici sous le n° GS915231.

La première séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°23.

La séquence SEQ ID N°23 a une longueur de 2764 nucléotides.

Cette séquence comprend un cadre ouvert de lecture (ORF)
15 partiel allant du nucléotide en position 3 au nucléotide en position 1220 de la séquence SEQ ID N°23.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

La deuxième séquence nucléique de l'ADNc correspondant à ce
20 transcrit constitue la séquence SEQ ID N° 24.

La séquence SEQ ID N° 24 a une longueur de 3228 nucléotides. Elle comprend un cadre ouvert de lecture (ORF) allant du nucléotide en position 37 au nucléotide en position 1304. La séquence codante débute au nucléotide en position 49 et se termine au nucléotide en position
25 1304 de la séquence SEQ ID N° 24. Le début du codon d'initiation de la traduction est localisé sur le nucléotide en position 49. Il existe un signal de polyadénylation débutant sur le nucléotide en position 3142 de la séquence SEQ ID N° 24.

Des homologies de séquence ont été observées entre la
30 séquence SEQ ID N° 24 et les séquences suivantes référencées dans la base de données GenBank (Version 116) :

- 100% d'homologie sur 217 pdb (positions [2704-2920]) matchant avec les positions [1-217] de la sequence g4884337 (AL050130) Homo sapiens mRNA; cDNA DKFZp586H051 (from clone DKFZp586H051). Length = 1795; deposee le 18-FEB-2000 ; Direct
5 Submission ; Submitted (15-MAY-1999) MIPS, Am Klopferspitz 18a, D-82152, Martinsried, GERMANY,
- 4 fragments d'homologie : 100% sur 393 pdb [2773-3165]; 100% sur 153 pdb [913-1065]; 100% sur 111 pdb [1083-1193]; 84% sur 84 pdb
10 [2341-2424] avec la sequence g6539402(AC016904) Homo sapiens clone RP11-307P9, *** SEQUENCING IN PROGRESS ***, 36 unordered pieces. Length = 203456;
- Differents fragments d'homologie : 83% sur 258 pdb [529-786]; 89%
15 sur 66pdb [1195-1260]; 80% sur 185pdb [91-275] avec la sequence g5305227 (AF029260) Gallus gallus transcription factor RelB (relb) mRNA, complete cds. Length = 2851

Les analyses d'expression du transcrit de séquence SEQ ID N°23
20 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS915231 était exprimé dans le cerveau foetal.

De plus, une analyse de l'expression du transcrit par Northern blot, selon le protocole décrit de l'Exemple 1, à l'aide de la sonde de
25 séquence SEQ ID N°91, a révélé la présence de transcrits dans le blot commercialisé par la Société Clontech (Ref. N° 7759-1).

La taille des transcrits détectés avec la sonde de séquence SEQ ID N°91 est respectivement de :

- 1,3 kb, 2 kb, 4 kb, 4,4 kb et 7,5 kb dans le cœur, et le muscle
30 squelettique :

- 1,3 kb, 2 kb, 4 kb et 4,4 kb dans le foie et le rein ;
- 1,3 kb, 2 kb, 4,4 kb et 7,5 kb dans le cerveau ;
- 1,3 kb, 2 kb et 4,4 kb dans le pancréas.

5 Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

10

Polypeptide codé par l'acide nucléique de séquence SEQ ID NO 23

Le cadre ouvert de lecture partiel de la séquence d'acides nucléiques SEQ ID N°23 code potentiellement pour un polypeptide de 406 aminoacides de longueur constituant la séquence SEQ ID N°139.

15

Homologies observées au niveau de la séquence protéique :

- 51% d'homologie avec des séquences de Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999): gi:2731377 est la traduction protéique de la séquence gb|AAB93456.1| (U28739) similaire à l'alcool déshydrogénase/ribitol déshydrogénase de *C. elegans* et des séquences de SP-TrEMBL (SP-TrEMBL, version 7, novembre 1998):Q09979 sur 401 aa (position 1-401 aa).
- 25 • 30% d'homologie sur 164 aa (position 47-205 aa) avec les séquences de Protein Data Bank: : gi|1827713| correspondant à la protéine 1AHI|A Chain A, 7 Alpha-Hydroxysteroid Dehydrogenase Complexed With NADH And 7-Oxo Glycochenodeoxycholic Acid et avec
- 30

gi|1827714| correspondant à la protéine 1AHI|B Chain B,
 7 Alpha-Hydroxysteroid Dehydrogenase Complexed
 With Nadh And 7-Oxo Glycochenodeoxycholic Acid et
 avec gi|1827715| correspondant à la protéine 1AHH|A
 Chain A, 7 Alpha-Hydroxysteroid Dehydrogenase
 Complexed With Nad⁺ et avec gi|1827716|
 correspondant à la protéine 1AHH|B Chain B, 7 Alpha-
 Hydroxysteroid Dehydrogenase Complexed With Nad⁺
 et avec gi|1943533| correspondant à la protéine 1FMC|A
 Chain A, 7-Alpha-Hydroxysteroid Dehydrogenase
 Complex With Nadh And 7-Oxo Glycochenodeoxycholic
 Acid et avec gi|1943534| correspondant à la protéine
 1FMC|B Chain B, 7-Alpha-Hydroxysteroid
 Dehydrogenase Complex With Nadh And 7-Oxo
 Glycochenodeoxycholic Acid

- 29% d'homologie sur 131 aa (position 53-183 aa) avec
 des séquences de Protein Data Bank: gi|2624497|
 correspondant à la protéine *Cis-Biphenyl-2,3-
 Dihydrodiol-2,3-Dehydrogenase* de *Pseudomonas Sp.*
Lb400
- 27% d'homologie sur 194 aa (position 3-196 aa) avec
 des séquences de la base PIR: (SEQUENCES PIR NON
 REDONDANTES, VERSION 57): PIR: (Séquences PIR
 non redondantes, version 57)|D69930| correspondant à
 la protéine *3-oxoacyl- acyl-carrier protein reductase
 homolog yoxD* de - *Bacillus subtilis* et avec des
 séquences de Swissprot (version 36, dernière remise à
 jour du 3 mai 1999): sp|P14802| correspondant à la

protéine YOXD_BACSU - oxidoreductase hypothétique
présente dans la région RTP-PELB (ORF238)

- 5 • 52% d'homologie sur 44 aa (position 353-396 aa) avec des séquences de PRODOM: (domaines homologues détectés dans Swissprot, version 34.2, novembre 1997) (homologous domains detected in Swissprot (version 36, dernière remise à jour du 3 mai 1999)): 2675 p34.2 (11)

10 NLTP(5) DHB4(3) PX18(2) correspondant à la protéine LIPID-TRANSFER STEROL CARRIER SCP-2 NONSPECIFIC PRECURSOR ESTRADIOL BETA-DEHYDROGENASE 17-BETA-HYDROXYSTEROID
- 15 • 27% d'homologie sur 167 aa (position 233-399 aa) avec des séquences de Swissprot (version 36, dernière remise à jour du 3 mai 1999) : sp|P51659| correspondant à la protéine DHB4_HUMAN ESTRADIOL 17 BETA-DEHYDROGENASE 4 (EC 1.1.1.62) (17-BETA-HSD 4)

20 (17-BETA-HYDROXYSTEROID DEHYDROGENASE 4)

Polypeptide codé par l'acide nucléique de séquence SEQ ID N° 24

Le cadre ouvert de lecture (ORF) de la séquence d'acide nucléique SEQ ID N° 24 code potentiellement pour un polypeptide de

25 422 acides aminés de longueur constituant la séquence SEQ ID N° 140.

Les homologies de séquences suivantes ont été retrouvées :

Homologies observées au niveau de la séquence protéique :

- 30 • 51% d'homologie avec Genpept: gi:2731377 et SP-trEMBL:Q09979 sur 416 aa (position 11-417 aa). Genpept: gi:2731377 est la traduction protéique de la

séquence gb|AAB93456.1| (U28739) similaire à l'alcool
déshydrogénase/ribitol déshydrogénase de *C. elegans*

- 5
- 30% d'homologie sur 164 aa (position 47-205 aa) avec Protein Data Bank: gi|1827713| correspondant à la protéine 1AHI|A Chain A, 7 Alpha-Hydroxysteroid Dehydrogenase Complexed With Nadh And 7-Oxo Glycochenodeoxycholic Acid et avec gi|1827714| correspondant à la protéine 1AHI|B Chain B, 7 Alpha-Hydroxysteroid Dehydrogenase Complexed With Nadh And 7-Oxo Glycochenodeoxycholic Acid et avec gi|1827715| correspondant à la protéine 1AHH|A Chain A, 7 Alpha-Hydroxysteroid Dehydrogenase Complexed With Nad+ et avec gi|1827716| correspondant à la protéine 1AHH|B Chain B, 7 Alpha-Hydroxysteroid Dehydrogenase Complexed With Nad+ et avec gi|1943533| correspondant à la protéine 1FMC|A Chain A, 7-Alpha-Hydroxysteroid Dehydrogenase Complex With Nadh And 7-Oxo Glycochenodeoxycholic Acid et avec gi|1943534| correspondant à la protéine 1FMC|B Chain B, 7-Alpha-Hydroxysteroid Dehydrogenase Complex With Nadh And 7-Oxo Glycochenodeoxycholic Acid
- 10
- 15
- 20
- 25
- 29% d'homologie sur 131 aa (position 53-183 aa) avec Protein Data Bank: gi|2624497| correspondant à la protéine Cis-Biphenyl-2,3-Dihydrodiol-2,3-Dehydrogenase From *Pseudomonas* Sp. Lb400

- 27% d'homologie sur 202 aa (position 6-212 aa) avec Non redundant PIR: pir[D69930] correspondant à la protéine 3-oxoacyl- acyl-carrier protein reductase homolog yoxD - Bacillus subtilis et avec Swissprot: sp[P14802] correspondant à la protéine YOXD_BACSU HYPOTHETICAL OXIDOREDUCTASE IN RTP-PELB INTERGENIC REGION (EC 1.-.-.-) (ORF238)
5
- 52% d'homologie sur 44 aa (position 353-396 aa) avec Prodom (homologous domains detected in Swissprot): 2675 p34.2 (11) NLTP(5) DHB4(3) PX18(2) correspondant à la protéine LIPID-TRANSFER STEROL CARRIER SCP-2 NONSPECIFIC PRECURSOR ESTRADIOL BETA-DEHYDROGENASE 17-BETA-HYDROXYSTEROID
10
15
- 27% d'homologie sur 183 aa (position 249-415 aa) avec Swissprot : sp[P51659] correspondant à la protéine DHB4_HUMAN ESTRADIOL 17 BETA-DEHYDROGENASE 4 (EC 1.1.1.62) (17-BETA-HSD 4) (17-BETA-HYDROXYSTEROID DEHYDROGENASE 4)
20
- 44% d'homologie sur 272 AA (positions 12-280 AA) avec la sequence g2072661 (Z95120) hypothetical protein Rv3224 [Mycobacterium tuberculosis] Length = 282
25
- Homologie avec des NONSPECIFIC LIPID-TRANSFER PROTEIN PRECURSOR (NSL-TP):

35 % sur 106 AA ([318-417]) avec la sequence
SP:NLTP_BOVIN NONSPECIFIC LIPID-TRANSFER
PROTEIN (STEROL CARRIER PROTEIN 2) (SCP-2).
Length = 121

5 et 30% sur 133 AA [294-417] avec la sequence
SP:NLTP_RAT NONSPECIFIC LIPID-TRANSFER
PROTEIN PRECURSOR (NSL-TP) (STEROL CARRIER
PROTEIN 2) (SCP-2) (STEROL CARRIER PROTEIN X)
(SCP-X) (SCPX).Length = 547

10

Fonction putative :

L'ORF de 422 acides aminés possède des homologies avec différentes
protéines enzymatiques putatives impliquées dans les mécanismes de
deshydrogénation des steroles de plusieurs organismes: homme, souris,
15 E. coli, S. cerevisiae, C. elegans. De plus nous avons pu mettre en
évidence une homologie avec une séquence de la protéine SCP-2
impliquée dans le transport intracellulaire des lipides. De ce fait, le gène
GS15231 code pour une protéine d'intérêt de par sa fonction possible
dans le défaut de trafic intracellulaire du cholestérol chez les patients
20 Tangier/FHD, ce qui renforce l'intérêt de l'étude du gène GS15231.
De plus le gène GS15231 est par sa localisation un gène candidat
positionnel pour l'étude et la caractérisation du défaut génétique observé
dans la maladie de Tangier ou le défaut familial en HDL.

Les polypeptides de séquences SEQ ID N° 139 et 140 sont
25 susceptibles d'intervenir dans la régulation du flux du cholestérol, et plus
particulièrement de la maladie de Tangier, dans des déficiences
familiales en HDL, ou encore dans une maladie génétiquement liée au
locus 9q31-34 du chromosome 9.

30

Gène GS915528.***Acide(s) nucléique(s)***

Il a été isolé selon l'invention deux ARNs messagers correspondant à un transcrit du gène désigné ici sous le n° GS915528.

5 La première séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°25.

La séquence SEQ ID N°25 a une longueur de 3106 nucléotides.

Cette séquence comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 1 au nucléotide en position 1272
10 de la séquence SEQ ID N°25.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

La deuxième séquence nucléique de l'ADNC correspondant à ce transcrit constitue la séquence SEQ ID N° 26.

15 La séquence SEQ ID N° 26 a une longueur de 3313 nucléotides. Elle comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 3 au nucléotide en position 1370, qui correspond aussi à la séquence codante. Un signal de polyadénylation débute au nucléotide en position 3280 de la séquence SEQ ID N° 26.

20 Les homologues de séquence des la SEQ ID N° 26 ont été observées avec des séquences référencées dans la base de données GenBank (Version 116) :

- 99% d'identité nucléique sur 2755pb (position 119-2873) avec
25 g7020444 AK000388 Homo sapiens cDNA FLJ20381 fis, clone KAIA2329 Length = 2970 pb

- 99% d'identité nucléique avec un BAC en cours de séquençage
g6514007 AC013568 Homo sapiens clone RP11-1B9, WORKING
30 DRAFT SEQUENCE 10 fragments non ordonnés

Les analyses d'expression du transcrit de séquence SEQ ID N°25 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS915528 était exprimé dans le cerveau fœtal, le foie, la prostate, le cœur, le placenta, l'utérus, le testicule et le cerveau.

De plus, une analyse de l'expression du transcrit par Northern blot, selon le protocole décrit de l'Exemple 1, à l'aide de la sonde de séquence SEQ ID N°92, a révélé la présence de transcrits dans le blot commercialisé par la Société Clontech (Ref. N° 7759-1).

La taille des transcrits détectés avec la sonde de séquence SEQ ID N°92 est respectivement de :

- 1,9 kb, 3,2 kb et 3,8 kb dans le pancréas ;
- 1 kb, 1,9 kb et 3,8 kb dans le cœur ;
- 1 kb, 1,9 kb et 3,2 kb dans le foie ;
- 1 kb et 1,9 kb dans le rein ;
- 1,9 kb dans le muscle squelettique et le cerveau.

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL.

Polypeptide codé par l'acide nucléique de séquence SEQ ID NO 25

Le cadre ouvert de lecture de la séquence d'acides nucléiques SEQ ID N°25 code potentiellement pour un polypeptide de 424 aminoacides de longueur constituant la séquence SEQ ID N°141.

Homologies observées au niveau de la séquence protéique :

Cette ORF de 424 aa possède des homologies de type motif avec des domaines conservés de différentes protéines tyrosine phosphatase dans différentes bases de données protéiques telles que :

- 5
- 49% d'homologie sur 364 aa (position 3-366 aa) avec des séquences de SP-TrEMBL (SP-TrEMBL, version 7, novembre 1998): sp|P91433|P91433 codé par l'ADNc YK65E9.3 de *C. elegans* et avec des séquences de Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999): gi|1708767| contenant un
- 10
- domaine trouvé dans la bande 4.1, ezrin, moesin, radixin et talin chez *C. elegans*
- 15
- 44% d'homologie sur 322 aa (position 1-322 aa) sur des séquences de SP-TrEMBL (SP-TrEMBL, version 7, novembre 1998): sp|O43491|O43491 PROTEIN 4.1-G
- 20
- 43% d'homologie sur 227 aa (position 7-233 aa) avec des séquences de PRODOM: (domaines homologues détectés dans Swissprot, version 34.2, novembre 1997): 894 p34.2 (29) MOES(4) RAD(3) EZRI(3) correspondant au domaine protéine tyrosine phosphatase des protéines MOESIN BAND P81 VILLIN-2 EZRIN RADIXIN impliquées dans l'organisation membranaire.
- 25
- 42% d'homologie sur 313 aa (position 9-321 aa) sur des séquences de Swissprot (version 36, dernière remise à jour du 3 mai 1999): sp|P29074|PTN4_HUMAN - protéine tyrosine phosphatase humaine MEG1 (EC 3.1.3.48).

30

- 41% d'homologie sur 320 aa (position 7-326 aa) avec des séquences de PIR: (SEQUENCES PIR NON REDONDANTES, VERSION 57) non-redondante: PIR: (Séquences PIR non redondantes, version 57)|S|JU0188
5 - protéine band 4.1 appartenant à la super-famille des protéines tyrosine phosphatase de souris. et des séquences de Swissprot (version 36, dernière remise à jour du 3 mai 1999): sp|P52963|NBL4_mouse - protéine NBL4 et des séquences de SP-TrEMBL (SP-TrEMBL, 10 version 7, novembre 1998): sp|O57457| protéine similaire à la protéine band 4.1 et des séquences de Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999): gi|466548| - protéine NBL4 de mus musculus
15
- 41% d'homologie sur 314 aa (position 9-322 aa) sur des séquences de Swissprot (version 36, dernière remise à jour du 3 mai 1999) : sp|P11171|41_HUMANPROTEIN 4.1 (BAND 4.1) (P4.1)
20

Polypeptide codé par l'acide nucléique de séquence SEQ ID N° 26

Le cadre ouvert de lecture de la séquence d'acide nucléique SEQ ID N° 26 code potentiellement pour un polypeptide de 455 acides aminés de longueur constituant la séquence SEQ ID N° 142.

25

Homologies observées au niveau de la séquence protéique :

Cette ORF de 455 aa possède des homologies de type motif avec des domaines conservés de différentes protéines tyrosine phosphatase dans différentes bases de données protéiques telle que .

- 50% d'homologie sur 374 aa (position 24-397 aa) avec
SP-trEMBL: sp|P91433|P91433 CODED FOR BY C.
ELEGANS CDNA YK65E9.3 et avec Genpept115:
gi|1708767| U80955 contenant un domaine trouvé dans
la bande 4.1, ezrin, moesin, radixin et talin chez c.
elegans
- 45% d'homologie sur 333 aa (position 21-353 aa) sur
SP-trEMBL: sp|O43491|O43491 PROTEIN 4.1-G.
- 43% d'homologie sur 227 aa (position 7-233 aa) avec
Prodom: 894 p34.2 (29) MOES(4) RADI(3) EZRI(3)
correspondant à la PROTEIN PHOSPHATASE
PROTEIN-TYROSINE MOESIN BAND P81 VILLIN-2
EZRIN RADIXIN MEMBRANE-ORGANIZING
- 42% d'homologie sur 329 aa (position 24-352 aa) sur
Swissprot38: sp|P29074|PTN4_HUMAN PROTEIN-
TYROSINE PHOSPHATASE MEG1 (EC 3.1.3.48)
(PTPASE-MEG1) (MEG).
- 42% d'homologie sur 335 aa (position 23-357aa) avec
PIR non-redondante: pir|S|JU0188 band 4.1 protein
tyrosine-phosphatase superfamily member protein de
souris et Swissprot: sp|P52963|NBL4_mouse NBL4
protein et SP-trEMBL: sp|O57457| band 4.1-like protein
4 et Genpept: gi|466548| NBL4 mus musculus protein

- 41% d'homologie sur 332 aa (position 22-353 aa) sur Swissprot : sp|P11171|41_HUMAN PROTEIN 4.1 (BAND 4.1) (P4.1).

5 Fonction putative :

Le gène GS915528 présente un intérêt pour l'étude fonctionnelle de la maladie de Tangier et du FHD en ce sens qu'il s'agit à la fois d'un candidat positionnel et d'un gène codant pour une nouvelle protéine dont les fonctions putatives (tyrosine phosphatase) pourraient jouer un rôle
10 dans la cascade de signalisation intracellulaire liée à l'efflux du cholestérol intracellulaire médié par les particules HDL, impliqué dans la maladie de Tangier et la déficience familiale en HDL (FHD).

Les polypeptides de séquences SEQ ID N° 141 et 142 sont susceptibles d'intervenir dans la régulation du flux du cholestérol, et plus
15 particulièrement de la maladie de Tangier, dans des déficiences familiales en HDL, ou encore dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

20 **Gène GS99817**

Acide(s) nucléique(s)

Il a été isolé selon l'invention deux ARNs messagers correspondant à un transcrit du gène désigné ici sous le n° GS99817.

La première séquence nucléique de l'ADNc correspondant à ce
25 transcrit constitue la séquence SEQ ID N°27.

La séquence SEQ ID N°27 a une longueur de 1539 nucléotides.

Cette séquence comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotidé en position 3 au nucléotide en position 698 de la séquence SEQ ID N°27.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

La seconde séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N° 28.

5 La séquence SEQ ID N° 28 a une longueur de 3404 nucléotides. Elle comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 1 au nucléotide en position 792 de la séquence SEQ ID N° 28.

Des homologues de séquence de la SEQ ID N° 28 ont été
10 retrouvées avec des séquences référencées dans la base de données GenBank (Version 116) :

- Identité de 97% sur 380 pb avec un BAC END CIT-HSP-2166G6.TR
CIT-HSP Homo sapiens genomic clone 2166G6, genomic survey
sequence Length = 380gi|2975337|gb|B93000.1|B93000[2975337]

15 - Identité de 100% sur 315 pb avec un BAC END HS_2166_A2_D03_MR
CIT Approved Human Genomic Sperm Library D Homo sapiens genomic
clone Plate=2166 Col=6 Row=G, genomic survey sequence Length =
316 gi|3480271|gb|AQ104915.1|AQ104915[3480271]

20

Les analyses d'expression du transcrit de séquence SEQ ID N°27 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS99817 était exprimé dans le cerveau foetal.

25 De plus, une analyse de l'expression du transcrit par Northern blot, selon le protocole décrit de l'Exemple 1, à l'aide de la sonde de séquence SEQ ID N°93, a révélé la présence de transcrits dans le blot commercialisé par la Société Clontech (Ref. N° 7759-1).

La taille des transcrits détectés avec la sonde de séquence SEQ
30 ID N°93 est respectivement de :

- 1,5 kb, 2 kb et 4,4 kb dans le cœur et le cerveau ;
- .2 kb et 4,4 kb dans le pancréas ;
- 1,5 kb at 4,4 kb dans le rein et le muscle squelettique.

- 5 Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

10

Polypeptide codé par l'acide nucléique de séquence SEQ ID NO 27

Le cadre ouvert de lecture de la séquence d'acide nucléique SEQ ID N°27 code potentiellement pour un polypeptide de 232 aminoacides de longueur constituant la séquence SEQ ID N°143.

15

- Homologies observées au niveau de la séquence protéique :
27% d'homologie sur 211 aa (position 11-221 aa) avec des séquences de Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999): gi|3876730| et des séquences de TrEMBL (SP-TrEMBL, version 7, novembre 1998): sp|Q20021| correspondant à la traduction de
20 la séquence d'un cosmide de nématode F35C11.4 (Caenorhabditis elegans).

Polypeptide codé par l'acide nucléique de séquence SEQ ID N° 28

- Le cadre ouvert de lecture de la séquence d'acide nucléique SEQ
25 ID N° 28 code potentiellement pour un polypeptide de 263 acides aminés de longueur constituant la séquence SEQ ID N° 144.

Il a été observé des homologies entre la séquence SEQ ID N° 144 et les séquences suivantes :

Homologies observées au niveau de la séquence protéique:

28% d'homologie sur 255 aa (position 1-255 aa) avec Genpept: gi|3876730| et trEMBL: sp|Q20021| correspondant à la traduction de la séquence d'un cosmide de nématode F35C11,4 (Caenorhabditis elegans).

5

Fonction putative :

Ce gène est un candidat pour l'étude de la maladie de Tangier/FHD dû fait de la localisation chromosomique.

Les polypeptides de séquences SEQ ID N° 143 et 144 sont susceptibles d'intervenir dans la régulation du flux du cholestérol, et plus particulièrement de la maladie de Tangier, dans des déficiences familiales en HDL, ou encore dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

15

Gène GS916229.

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messenger correspondant à un transcrit du gène désigné ici sous le n° GS916229.

La séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°29.

La séquence SEQ ID N°29 a une longueur de 792 nucléotides.

Cette séquence comprend un cadre ouvert de lecture (ORF) allant du nucléotide en position 1 au nucléotide en position 203 de la séquence SEQ ID N°29.

25

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

Les analyses d'expression du transcrit de séquence SEQ ID N°29 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces

30

analyses effectuées à partir d'ARN polyA+de différents tissus ont permis de montrer que le gène GS916229 était exprimé dans le cerveau foetal, le foie, le cerveau, le cœur, la prostate, le placenta, le foie foetal, l'utérus, le testicule et le rein.

5

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

10

Polypeptide codé par l'acide nucléique de séquence SEQ ID NO 29

Le cadre ouvert de lecture de la séquence d'acides nucléiques SEQ ID N°29 code potentiellement pour un polypeptide de 68 aminoacides de longueur constituant la séquence SEQ ID N°145.

15

Aucune homologie significative avec BLAST n'a été retrouvée avec les séquences référencées dans les bases de données Swissprot (version 36, dernière remise à jour du 3 mai 1999) et PRODOM: (domaines homologues détectés dans Swissprot, version 34.2, novembre 1997).

20

Le polypeptide de séquence SEQ ID N° 145 est susceptible d'intervenir dans la régulation du flux du cholestérol, et plus particulièrement de la maladie de Tangier, dans des déficiences familiales en HDL, ou encore dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

25

Gène GS92544

30

Acide(s) nucléique(s)

Il a été isolé selon l'invention trois ARN messagers correspondant respectivement à un transcrit long et deux transcrits courts du gène désigné ici sous le n° GS92544.

La séquence nucléique de l'ADNc correspondant au transcrit long
5 constitue la séquence SEQ ID N°30.

La séquence SEQ ID N°30 a une longueur de 2733 nucléotides.

Cette séquence comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 1 au nucléotide en position 2160 de la séquence SEQ ID N°30.

10

La première séquence nucléique de l'ADNc correspondant au transcrit court constitue la séquence SEQ ID N°31.

La séquence SEQ ID N°31 a une longueur de 2694 nucléotides.

Cette séquence comprend un cadre ouvert de lecture (ORF)
15 partiel allant du nucléotide en position 1 au nucléotide en position 2121 de la séquence SEQ ID N°31.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

La seconde séquence nucléique de l'ADNc correspondant au
20 second transcrit court constitue la séquence SEQ ID N° 32.

La séquence SEQ ID N° 32 a une longueur de 2765 nucléotides. Elle comprend un cadre ouvert de lecture complet allant du nucléotide en position 56 au nucléotide en position 2287 de la séquence SEQ ID N° 32. La séquence codante débute au nucléotide en position 410 et se
25 termine au nucléotide en position 2160 de la séquence SEQ ID N° 32.. Le codon d'initiation de la traduction débute sur le nucléotide en position 410 de la séquence SEQ ID N° 32.

La séquence SEQ ID N° 32 possède des homologues avec les séquences référencées dans la base de données GenBank (Version
30 116) suivantes :

- 100% d'homologie sur 2419 pdb (positions [485-2903]) matchant avec les positions [1-2419] de la sequence g6807990 (AL137432) Homo sapiens mRNA; cDNA DKFZp761E1824 (from clone DKFZp761E1824); partial cds , length=2438 pdb , deposee le 18-
5 FEB-2000, sur Chromosome 9, et annotee " similaire au CR2 receptor "
- 97 % d'homologie sur 157 pdb (positions [1271-1427])avec les positions [431-277] de g3590696 (AQ192074)
10 HS_3228_B2_H11_T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3228 Col=22 Row=P, genomic survey sequence. Length = 513
- Plusieurs fragments de 99% à 100% d'identité avec la sequence
15 gi|6982613| |AL138756 : Homo sapiens chromosome 9 clone RP11-4O1 map q31.3-33.1, *** SEQUENCING IN PROGRESS ***, 41 unordered pieces
- Plusieurs fragments de 100% d'identité avec la sequence g7230026
20 (AC010824) Homo sapiens clone RP11-5A23, *** SEQUENCING IN PROGRESS ***, 32 unordered pieces. Length = 162010

Les analyses d'expression du transcrit long et du transcrit court ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces
25 analyses effectuées à partir d'ARN polyA+de différents tissus ont permis de montrer que le gène GS92544 était exprimé dans le cerveau foetal.

De plus, une analyse de l'expression de ces transcrits par Northern blot, à l'aide respectivement des sondes de séquence SEQ ID N°94, a révélé la présence de transcrits dans le blot commercialisé par la
30 Société Clontech (Ref. N° 7759-1).

La taille des transcrits détectés avec la sonde de séquence SEQ ID N°94 est respectivement de 4 kb et de 6 kb dans le placenta.

5 Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

10 ***Polypeptides codés par les acides nucléiques de séquences SEQ ID NO 30 et 31***

Le cadre ouvert de lecture de la séquence d'acides nucléiques SEQ ID N°30 (transcrit long) code potentiellement pour un polypeptide de 720 aminoacides de longueur constituant la séquence SEQ ID N°146.

15 Le cadre ouvert de lecture de la séquence d'acides nucléiques SEQ ID N°31 (transcrit court) code potentiellement pour un polypeptide de 707 aminoacides de longueur constituant la séquence SEQ ID N°147.

Homologies observées au niveau de la séquence protéique:

20 Cette séquence possède des homologies significatives des positions 139-194 aa / 199 - 254 aa avec des domaines conservés sushi.HMM de la base de données Pfam (HMMER 2.0 (June 1998).

Cette ORF possède une annotation dans Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999) et TrEMBL (SP-TrEMBL, version 7, novembre 1998) avec la traduction BlastX de
25 nombreuses séquences telles que :

- 42% d'homologie sur 115 aa (position 2-116 aa) avec des séquences de gi|340164| précurseur humain de l'uromoduline et gi|340166| uromoduline [Homo sapiens]

sur Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999)

- 5 • 37% d'homologie sur 141 a (position 2-142 aa) avec des séquences de SP-TrEMBL (SP-TrEMBL, version 7, novembre 1998): sp|P87363|P87363 un fragment de la FIBRILLINE-1.
- 10 • 30% d'homologie sur 234 aa (position 7-240 aa) avec des séquences de Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999): gi|306746| et gi|1335064| fibrilline humaine.
- 15 • 30% d'homologie sur 194 aa (position 8-201 aa) avec des séquences de SP-TrEMBL (SP-TrEMBL, version 7, novembre 1998): sp|O35806|O35806 *LATENT TGF-BETA BINDING PROTEIN-2 LIKE PROTEIN*

Polypeptide codé par l'acide nucléique de séquence SEQ ID N° 32

20 Le cadre ouvert de lecture de la séquence d'acide nucléique SEQ ID N° 32 code potentiellement pour un polypeptide de 713 acides aminés de longueur constituant la séquence SEQ ID N° 148.

Des homologies de la séquence SEQ ID N° 148 ont été observées avec des séquences référencées dans des bases de données. Les homologies sont les suivantes :

- 25 • homologie avec Genpept: gi|340164| uromodulin precursor [Homo sapiens] et Genpept: gi|340166| uromodulin [Homo sapiens]
- 30 • homologie avec SP-trEMBL: sp|P87363|P87363 FIBRILLIN-1 (FRAGMENT).

- homologie avec Genpept: gi|306746| et gi|1335064| fibrillin [Homo sapiens]
- homologie avec SP-trEMBL: sp|O35806|O35806 LATENT TGF-BETA BINDING PROTEIN-2 LIKE PROTEIN
- homologie avec g784994 (X81479) EMR1 [Homo sapiens] Length = 886
- homologie avec g4379069 (X94630) seven-span transmembrane protein [Homo sapiens]
- Autres homologies avec des protéines telles que : EMR1, CD97, fibulins, complement receptor...

La séquence SEQ ID N° 148 possède des motifs caractéristiques tels que 3 domaines EGF dont 2 calcium binding, un site de tyrosine phosphatase, un domaine hydrophobe en N-terminal ; Nombreux sites de glycosylation ; 2 camp sites de phosphorylation ; 2sites Asp hydroxylation.

Fonction putative :

De par son homologie avec l'uromoduline, le produit du gène GS92544 est putativement une protéine associée à la membrane comme l'uromoduline qui est une protéine ancrée au moyen d'un glycosylphosphatidylinositol (GPI). Ces résultats basés sur les études d'homologie de séquence partielles en acides aminés suggèrent que le produit du gène GS92544 pourrait être associé à la membrane par une liaison aux lipides membranaires et donc pourrait être associé au déficit d'efflux du cholestérol cellulaire observé chez les patients FHD ou Tangier.

Les polypeptides de séquences SEQ ID N° 146, 147 et 148 sont susceptibles d'intervenir dans la régulation du flux du cholestérol, et plus

particulièrement de la maladie de Tangier, dans des déficiences familiales en HDL, ou encore dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

5 **Gène GS930824**

Acide(s) nucléique(s)

Il a été isolé selon l'invention deux ARNs messagers correspondant à un transcrit du gène désigné ici sous le n° GS930824.

La première séquence nucléique de l'ADNc correspondant à ce
10 transcrit constitue la séquence SEQ ID N°33.

La séquence SEQ ID N°33 a une longueur de 4745 nucléotides.

Cette séquence comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 2 au nucléotide en position 514 de la séquence SEQ ID N°33.

15 Homologies observées au niveau de la séquence nucléotidique:
90% d'homologie avec Genbank :AF115435, syntaxine 17 de rat sur 510 pb (position 22-531 pb).

La seconde séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N° 34 .

20 La séquence SEQ ID N° 34 a une longueur de 5241 nucléotides. Elle comprend un cadre ouvert de lecture (ORF) complet allant du nucléotide en position 57 au nucléotide en position 1013 de la séquence SEQ ID N°34. La séquence codante débute au nucléotide en position 105 et se termine au nucléotide en position 1013 de la séquence SEQ ID
25 N° 34. Le codon d'initiation de la traduction débute sur le nucléotide en position 105 de la séquence SEQ ID N°34.

La séquence SEQ ID N° 34 possède des homologies avec des séquences référencées dans la base de données GenBank (Version 116). Il s'agit des homologies suivantes :

30

- 90% d'homologie avec Genbank :AF115435, syntaxine 17 de rat sur 510 pb (position 22-531 pb).

- 92 % d'homologie sur 475 pdb (positions [540-1036]) et 84% sur 406 pdb (positions [102-507]) avec la sequence g4206160 . (AF115435) Rattus norvegicus syntaxin 17 mRNA, complete cds.Length = 1678;
5
- 98% d'homologie sur 431 pdb (positions [1899-2329]) et 91% sur 62 pdb (positions [1819-1880]) avec la sequence g4652677 (AQ474416) CITBI-E1-258819.TF CITBI-E1 Homo sapiens genomic clone 258819, genomic survey sequence.Length = 525;
10
- 99% d'homologie sur 331 pdb (positions [6394-6724]) avec la sequence g2929043 (B87911) RPCI11-30N20.TP RPCI-11 Homo sapiens genomic clone RPCI-11-30N20, genomic survey sequence.Length = 425;
15
- 99% d'homologie sur 2662 pdb (positions [20-2681]) avec la sequence g7020892 (AK000658) Homo sapiens cDNA FLJ20651 fis, clone KAT01814.Length = 2678; Submitted (15-FEB-2000) to the DDBJ/EMBL/GenBank databases, projet NEDO.
20

Les analyses d'expression du transcrit de séquence SEQ ID N°33 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+de différents tissus ont permis
25 de montrer que le gène GS930824 était exprimé dans le cerveau foetal, le foie, le cerveau, le cœur, la prostate, le placenta, le foie foetal, l'utérus, le testicule, le rein et le muscle squelettique.

De plus, une analyse de l'expression du transcrit par Northern blot, selon le protocole décrit de l'Exemple 1, à l'aide respectivement des
30 sondes de séquence SEQ ID N° 95 et 96, a révélé la présence de

transcrits dans le blot commercialisé par la Société Clontech (Ref. N° 7759-1).

La taille des transcrits détectés avec la sonde de séquence SEQ ID N°95 est respectivement de :

- 5 - 1,1 kb, 1,6 kb, 2,6 kb, 4,9 kb. et 7 kb dans le pancréas, le rein, le muscle squelettique, le foie, le cerveau et le cœur ;
- 1,6 kb, 2,6 kb, 4,9 kb. et 7 kb dans le poumon et le placenta.

La taille des transcrits détectés avec la sonde de séquence SEQ ID N°96 est respectivement de :

- 10 - 1,35 kb, 2,4 kb, 3,5 kb et 10 kb dans le placenta ;
- 1,35 kb et 2,4 kb dans le pancréas, le rein et le foie ;
- 1,35 kb dans le poumon ;
- 2,4 kb dans le muscle squelettique, le cerveau et le cœur.

15

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL.

20

Polypeptide codé par l'acide nucléique de séquence SEQ ID NO 33

Le cadre ouvert de lecture de la séquence d'acides nucléiques SEQ ID N°33 code potentiellement pour un polypeptide de 170 aminoacides de longueur constituant la séquence SEQ ID N°149.

25

Aucune homologie significative avec BLAST n'a été retrouvée avec les séquences référencées dans les bases de données Swissprot (version 36, dernière remise à jour du 3 mai 1999) et PRODOM: (domaines homologues détectés dans Swissprot, version 34.2, novembre 1997).

30

Homologies observées au niveau de la séquence protéique:

72% d'homologie avec des séquences de Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999) :gi4206161 sur 170 acides aminés (position 1-170 aa) gi4206161 est la traduction protéique du gène codant pour la syntaxine 17 de rat..Genpept:
5 (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999) :

Polypeptide codé par l'acide nucléique de séquence SEQ ID N° 34

Le cadre ouvert de lecture de la séquence SEQ ID N° 34 code
10 potentiellement pour un polypeptide de 318 acides aminés constituant la séquence SEQ ID N° 150.

Les homologies suivantes ont été observées :

- 72% d'homologie avec Genpept :gi4206161 sur 170 acides aminés
15 (position 1-170 aa). Genpept :gi4206161 est la traduction protéique du gène codant pour la syntaxine 17 de rat.

Homologie protéique sur Genpept115, Trembl et PIR :

- 75% d'homologie sur 302 AA (positions [105-1010]) avec toute la
20 longueur la sequence traduite g4206161 (AF115435) syntaxin 17 [Rattus norvegicus]Length = 301; de la g4206161

Le polypeptide de séquence SEQ ID N° 150 possède un motif caractéristique d'une syntaxine entre les nucléotides en positions 1 et
25 243 du cadre de lecture ouvert.

Fonction putative :

Les polypeptides de séquence SEQ ID N°149 et 150 s'apparentent donc à une protéine de la famille des syntaxines impliquée
30 dans le transport vésiculaire intracellulaire. Ce mécanisme sous-tendu

par un adressage spécifique médié par ces protéines est d'intérêt par rapport à la maladie de Tangier/FHD qui est décrite par un défaut du mécanisme de transport et de translocation des pools de cholestérol intracellulaire vers les particules HDL accepteurs. Les polypeptides de
5 séquence SEQ ID N°149 et 150 sont donc susceptibles d'intervenir dans une étape importante impliquée dans le transport inverse du cholestérol par les HDL.

Les polypeptides de séquences SEQ ID N° 149 et 150 sont également susceptibles d'intervenir dans une maladie génétiquement
10 liée au locus 9q31-34 du chromosome 9.

Gène GS93382.

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messager correspondant à
15 un transcrit du gène désigné ici sous le n° GS93382.

La séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°35.

La séquence SEQ ID N°35 a une longueur de 3014 nucléotides.

Cette séquence comprend un cadre ouvert de lecture (ORF)
20 partiel allant du nucléotide en position 3 au nucléotide en position 371 de la séquence SEQ ID N°35.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

25 Les analyses d'expression du transcrit de séquence SEQ ID N°35 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS93382 était exprimé dans le cerveau foetal.

De plus, une analyse de l'expression du transcrit par Northern
30 blot, selon le protocole décrit de l'Exemple 1, à l'aide de la sonde de

séquence SEQ ID N°97, a révélé la présence de transcrits dans le blot commercialisé par la Société Clontech (Ref. N° 7759-1).

La taille des transcrits détectés avec la sonde de séquence SEQ ID N°97 est respectivement de :

- 5 - 2 kb et 7,5 kb dans le cerveau ;
- 2 kb dans le pancréas, le rein, le muscle squelettique, le foie et le cœur.

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

15 ***Polypeptide codé par l'acide nucléique de séquence SEQ ID NO 35***

Le cadre ouvert de lecture de la séquence d'acides nucléiques SEQ ID N°35 code potentiellement pour un polypeptide de 123 aminoacides de longueur constituant la séquence SEQ ID N°151.

Aucune homologie significative avec BLAST n'a été retrouvée avec les séquences référencées dans les bases de données Swissprot (version 36, dernière remise à jour du 3 mai 1999), PRODOM: (domaines homologues détectés dans Swissprot, version 34.2, novembre 1997), Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999), Swissprot (version 36, dernière remise à jour du 3 mai 1999), TrEMBL (SP-TrEMBL, version 7, novembre 1998), 25 PIR: (SEQUENCES PIR NON REDONDANTES, VERSION 57) et PDB: (PROTEIN DATA BANK, FEVRIER 1999).

Le polypeptide de séquence SEQ ID N° 151 est susceptible d'intervenir dans la régulation du flux du cholestérol, et plus 30 particulièrement de la maladie de Tangier, dans des déficiences

familiales en HDL, ou encore dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

5 **Gène GS946300.**

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS946300.

La séquence nucléique de l'ADNc correspondant à ce transcrit
10 constitue la séquence SEQ ID N°36.

La séquence SEQ ID N°36 a une longueur de 1575 nucléotides.

Cette séquence comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 3 au nucléotide en position 176 de la séquence SEQ ID N°36.

15 Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

Les analyses d'expression du transcrit de séquence SEQ ID N°36 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces
20 analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS946300 était exprimé dans le cerveau fœtal, le foie, le cerveau, le cœur, la prostate, le placenta, le foie fœtal, l'utérus, le testicule et le rein.

25 Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

30

Polypeptide codé par l'acide nucléique de séquence SEQ ID NO 36

Le cadre ouvert de lecture de la séquence d'acides nucléiques SEQ ID N°36 code potentiellement pour un polypeptide de 58 aminoacides de longueur constituant la séquence SEQ ID N°152.

5 Aucune homologie significative avec BLAST n'a été retrouvée avec les séquences référencées dans les bases de données Swissprot (version 36, dernière remise à jour du 3 mai 1999) et PRODOM: (domaines homologues détectés dans Swissprot, version 34.2, novembre 1997).

10 Le polypeptide de séquence SEQ ID N°152 est susceptible d'intervenir dans la régulation du flux de cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL.

15 Le polypeptide de séquence SEQ ID N°152 est donc susceptible d'intervenir dans une étape importante impliquée dans le transport inverse du cholestérol par les HDL.

Gène GS937345.

20 ***Acide(s) nucléique(s)***

Il a été isolé selon l'invention deux ARN messagers correspondant respectivement à un transcrit long et un transcrit court du gène désigné ici sous le n° GS937345.

25 Une première séquence nucléique de l'ADNc correspondant au transcrit long a été isolée et constitue la séquence SEQ ID N°37.

La séquence SEQ ID N°37 a une longueur de 1607 nucléotides.

Cette séquence comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 2 au nucléotide en position 109 de la séquence SEQ ID N°37.

A partir de la séquence SEQ ID N° 37, deux amorces de séquences respectives SEQ ID N° 104 et 105 ont été synthétisées et ont permis d'amplifier un ADNc à partir d'une banque d'ARNm polyA+ de différents tissus humains commercialisée par la société Clontech.

5 La séquence d'ADNc amplifiée à l'aide des amorces de séquence SEQ ID N° 104 et 105 constitue la séquence SEQ ID N° 38. La séquence SEQ ID N° 38 a une longueur de 1161 nucléotides.

Des homologies ont été retrouvées entre la séquence SEQ ID N° 38 et des séquences référencées dans la base de données GenBank
10 (Version 116). Les homologies sont les suivantes :

- 99% d'homologie sur 1096 pdb (positions [1-1093]) matchant avec les positions [8-1102] de la séquence

15 gi|6841231|gb|AF161409.1|AF161409[6841231]Homo sapiens
HSPC291 mRNA, partial cds. Length = 1102;
unpublished ;

- 99% sur 1025 pdb (positions [119-1148]) matchant avec les positions [8-1030] de la séquence
20 gi|6841235|gb|AF161411.1|AF161411[6841235] Homo sapiens
HSPC293 mRNA, partial cds. Length = 1045; unpublished ;

- 99% sur 1161 pdb (positions [1-1161]) matchant avec les positions
25 [43-1202] de la séquence
gi|7020861|dbj|AK000637.1|AK000637[7020861] Homo sapiens
cDNA FLJ20630 fis, clone KAT03874.Length = 1538; déposée le 22-
FEB-2000, projet NEDO ; unpublished .

- 92 % sur 1003 pdb (positions [1-1043]) et 94% sur 38 pdb (positions [1048-1085]) avec la sequence gb|AC021286.2|AC021286[6899766] Homo sapiens clone RP11-21H22, WORKING DRAFT SEQUENCE, 19 unordered pieces.
5 Length = 175143 ;

La séquence nucléique de l'ADNc correspondant au transcrit court constitue la séquence SEQ ID N°39.

La séquence SEQ ID N°39 a une longueur de 1332 nucléotides.

10

Aucune identité de séquence n'a été retrouvée pour les séquences respectives du transcrit long et du transcrit court lors d'une recherche dans la base de données GenBank (Version 110).

- 15 Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

20

Polypeptide codé par l'acide nucléique de séquence SEQ ID NO 37

Le cadre ouvert de lecture partiel du transcrit long de séquence d'acides nucléiques SEQ ID N°37 code potentiellement pour un polypeptide de 36 aminoacides de longueur constituant la séquence
25 SEQ ID N°153.

Homologies observées au niveau de la séquence nucléotidique:
41% d'identité sur 22 aa (position 6 à 29 aa) avec des séquences de Swissprot (version 36, dernière remise à jour du 3 mai 1999): sp|P23596|
PRTD_ERWCH Proteases secretion ATP- Binding protein PRTD.

30

Ce gène est un candidat pour l'étude de la maladie de Tangier/FHD du fait de la localisation chromosomique.

Gène GS99556

5 *Acide(s) nucléique(s)*

Il a été isolé selon l'invention deux ARN messagers correspondant respectivement à un transcrit long et à un transcrit court du gène désigné ici sous le n° GS99556.

La séquence nucléique de l'ADNc correspondant au transcrit long
10 constitue la séquence SEQ ID N°40.

La séquence SEQ ID N°40 a une longueur de 10419 nucléotides.

Cette séquence comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 2 au nucléotide en position 1954 de la séquence SEQ ID N°40.

15 Le codon d'initiation (ATG) débute au nucléotide en position 29 du transcrit long de séquence SEQ ID N°40.

La séquence nucléique de l'ADNc correspondant au transcrit court constitue la séquence SEQ ID N°41.

20 La séquence SEQ ID N°41 a une longueur de 1813 nucléotides.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

25 Les analyses d'expression du transcrit de séquence SEQ ID N°40 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS99556 était exprimé dans le cerveau fœtal, le foie, le cerveau, le cœur, la prostate, le placenta et le foie fœtal.
30

De plus, une analyse de l'expression du transcrit par Northern blot, selon le protocole décrit de l'Exemple 1, à l'aide respectivement des sondes de séquence SEQ ID N°98 et SEQ ID N°99, a révélé la présence de transcrits dans le blot commercialisé par la Société Clontech (Ref. N°
5 7759-1).

La taille des transcrits détectés avec la sonde de séquence SEQ ID N°98 est respectivement de :

- 2,6 kb , 4,2 kb, 5 kb et 10 kb dans le cerveau ;
- 2,6 kb et 5 kb dans le foie, le poumon, le placenta et le cœur ;
- 10 - 2,6 kb et 5 kb dans le rein ;
- 2,6 kb dans le muscle squelettique ;
- 5 kb dans le pancréas.

La taille des transcrits détectés avec la sonde de séquence SEQ
15 ID N°99 est respectivement de :

- 2,2 kb dans le foie ;
- 2,4 kb et 4,4 kb dans le cœur ;
- 9 kb dans le cerveau, le placenta, le rein, le pancréas et le poumon.

20 La taille des transcrits détectés avec la sonde de séquence SEQ ID N° 100 est respectivement de :

- 5 kb et 7 kb dans le placenta et le cœur ;
- 5 kb dans le cerveau, le rein et le pancréas.

25 Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Polypeptide codé par l'acide nucléique de séquence SEQ ID NO 40

Le cadre ouvert de lecture partiel de la séquence d'acides nucléiques SEQ ID N°40 code potentiellement pour un polypeptidé de
5 651 aminoacides de longueur constituant la séquence SEQ ID N°154.

Homologies observées au niveau de la séquence protéique :

Cette ORF possède des homologies avec des séquences de Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999) et de TrEMBL (SP-TrEMBL, version 7, novembre 1998)

10 (traduction BlastX des séquences suivantes) :

- 32% d'homologie sur 403 aa dans Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999) avec gi|4529890|
15 NG22 [Homo sapiens]

- 25% d'homologie sur 693 aa dans Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999) avec gi|3986770|
20 NG22 [Mus musculus]

24% d'homologie sur 683 aa dans Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le 7 mai 1999) avec gi|1072187| correspondant à l'ADNc CEESB82F de C elegans

25

- 24% d'homologie sur 683 aa avec TrEMBL (SP-TrEMBL, version 7, novembre 1998) sp|Q20026| codé par l'ADNc CEESB82F de C elegans

30

Fonction putative :

Ce gène est un candidat pour l'étude de la maladie de Tangier/FHD du fait de la localisation chromosomique.

Le polypeptide de séquence SEQ ID N° 154 est susceptible d'intervenir dans la régulation du flux du cholestérol, et plus particulièrement de la maladie de Tangier, dans des déficiences familiales en HDL, ou encore dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

10 **Gène GS96663.**

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messenger correspondant à un transcrit du gène désigné ici sous le n° GS96663. Six séquences nucléiques représentatives de ce transcrit ont été déterminées :

15 La première séquence nucléique partielle de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°42.

La séquence SEQ ID N°42 a une longueur de 1377 nucléotides.

La seconde séquence nucléique partielle de l'ADNc
20 correspondant à ce transcrit constitue la séquence SEQ ID N°43.

La séquence SEQ ID N°43 a une longueur de 452 nucléotides.

La troisième séquence nucléique partielle de l'ADNc
correspondant à ce transcrit constitue la séquence SEQ ID N°44.

25 La séquence SEQ ID N°44 a une longueur de 562 nucléotides.

La quatrième séquence nucléique partielle de l'ADNc
correspondant à ce transcrit constitue la séquence SEQ ID N°45.

La séquence SEQ ID N°45 a une longueur de 1766 nucléotides.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

A partir de la séquence SEQ ID N° 42, on a synthétisé deux amorces nucléotidiques, respectivement les amorces de séquences
5 SEQ ID N° 106 et 107.

A partir de la séquence SEQ ID N° 43, on a synthétisé une amorce nucléotidique de séquence SEQ ID N° 108.

A partir de la séquence SEQ ID N° 45, on a synthétisé deux amorces nucléotidiques, respectivement les amorces de séquences
10 SEQ ID N° 109 et 110.

Ces amorces ont permis d'amplifier une cinquième et une sixième séquence nucléotidique représentatives du transcrit du gène GS96663.

La cinquième séquence nucléique correspondant au transcrit du gène GS96663 constitue la séquence SEQ ID N° 46. La séquence
15 nucléique SEQ ID N° 46 a une longueur de 601 nucléotides.

La sixième séquence nucléique correspondant au transcrit du gène GS96663 constitue la séquence SEQ ID N° 47. La séquence SEQ ID N° 47 a une longueur de 3706 nucléotides. Elle comprend un cadre
20 ouvert de lecture partiel allant du nucléotide en position 1 au nucléotide en position 3202 de la séquence SEQ ID N° 47.

Des homologues ont été observées entre la séquence SEQ ID N° 47 et des séquences référencées dans la base de données GenBank (Version 116). Ces homologues sont les suivantes :

- 25
- 99 % d'homologie sur 2423 pdb (positions [1030-3451]) matchant avec les positions [1-2419] de la séquence gi|5102585|emb|AL079279.1|HST000009[5102585] Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 248114 . Length =
30 2450 ; /chromosome="9" /map="D9S176-D9S279" ; deposee le 14-JUN-1999 ; Unpublished.

- 98% sur 1623 pdb (positions [1946-3559]) matchant avec les positions [16-1638] de la sequence
g3012351 (I76197) Sequence 1 from patent US 5691147.Length =
1638 pdb ; deposee le 03-APR-1998 ; AUTHORS : Draetta,G. and
5 Gyuris,J. TITLE : CDK4 binding assay JOURNAL : Patent: US
5691147-A 1 25-NOV-1997;

- Differents fragments d'homologie de 99% a 100% sur 2372 pdb
10 (positions [1-2372]) et de 97% a 100% sur 1160 pdb (positions [2547-
3706]) avec la sequence
gi|7228016|emb|AL158158.3|AL158158[7228016] ; Homo sapiens
chromosome 9 clone RP11-427L11 map q31.2-32, ***
SEQUENCING IN PROGRESS *** , 37 unordered pieces ; deposee le
15 08-MAR-2000 ;

Les analyses d'expression du transcrit des séquences SEQ ID
N°42 à 47 ont été réalisées par RT PCR, comme décrit dans l'Exemple

- 20 1. Ces analyses effectuées à partir d'ARN polyA+de différents
tissus ont permis de montrer que le gène GSS96663 était exprimé dans
le cerveau foetal, le foie, le cerveau, le cœur, la prostate, le placenta, le
foie foetal, l'utérus, le testicule, le rein et le muscle squelettique.

- 25 Ce gène constitue un candidat positionnel causal d'une maladie
due à un dysfonctionnement du flux inverse du cholestérol, et plus
particulièrement de la maladie de Tangier ou encore des déficiences
familiales en HDL.

Polypeptide codé par l'acide nucléique de séquence SEQ ID N° 47

Le cadre ouvert de lecture partiel de la séquence d'acide nucléique SEQ ID N° 47 code potentiellement pour un polypeptide de 1066 acides aminés de longueur qui constitue la séquence SEQ ID N° 155..

Des homologies entre la séquence SEQ ID N° 155 et des séquences répertoriées dans les bases de données. Ces homologies sont les suivantes :

- 10 • 27% d'homologie sur 1068 AA avec des sequences de type " complement receptor " CR1 (g30186 ; g809019 ; g451303 ;g306680) et CR2 (g18192 ;g181940 ;g599776)
- 15 • Homologie de 24% sur environ 1000 AA avec des sequences "complement facteur H" ; sequences gi|31965|emb|CAA68704.1|[31965] (Y00716) factor H [Homo sapiens], PIR :NBHUH et PIR :NBMSH ; EM:Q14006 et EM :Q61408.
- 20 • Homologie de 25% sur 900 AA avec les sequences de P-selectin et E-selectin : sequences sp|p16581|lem2_human e-selectin precursor (endothelial leukocyte adhesion molecule 1) (elam-1) (leukocyte-endothelial cell adhesion molecule 2) (lecam2) (cd62e)

25 et sp|p16109|lem3_human p-selectin precursor (granule membrane protein 140) (gmp-140) (padgem) (cd62p) (leukocyte-endothelial cell adhesion molecule 3) (lecam3) et avec pir non-redondante: pir|s|a30359 p-selectin precursor - human..

- 24% d' homologie sur 637 AA avec la sequence g183391 (M25322) granule membrane protein-140 (GMP-140) precursor [Homo sapiens] , Length = 830
- 5
- Homologie avec la " cell adhesion molecule " trEMBL: sp|Q28290|Q28290 CELL ADHESION MOLECULE PRECURSOR (FRAGMENT).
- 10
- Differents fragments d'homologie de 25% a 29% sur la sequence " apolipoprotein H precursor - human PIR:NBHU Length = 345; apolipoprotein H precursor - human. avec une homologie de 29% sur 256 AA.
- 15
- Homologie avec des sequences " membrane cofacteur protein cofacteur " (PIR :S01896, PIR :I54479, PIR :A57278 et EM :P79138, EM :Q9Z0M4, EM :O19121) et EM :O62837 : membrane cofacteur proteine CD46.
- 20
- Differents fragments (de 400 a 500 AA) d'homologie de 25% a 27% avec la sequence PIR:T16833 hypothetical protein T07H6.5 - Caenorhabditis elegans retrouvee dans genpept (g1255889(U53344).
- 25

30

Fonction putative :

Le produit du gène GS96663, de par son homologie en acides aminés avec les précurseurs de la P-sélectine et de la E-sélectine humaines, s'apparente à cette classe de protéines membranaires. Son rôle dans l'efflux du cholestérol intracellulaire, médié par des protéines membranaires n'est pas exclu d'autant que le gène est localisé dans
5 l'intervalle génétique défini par clonage positionnel.

Le polypeptide de séquence SEQ ID N° 155 est susceptible d'intervenir dans la régulation du flux du cholestérol, et plus particulièrement de la maladie de Tangier, dans des déficiences
10 familiales en HDL, ou encore dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Gène GS941675.

Acide(s) nucléique(s)

15 Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS941675. Deux séquences nucléiques sont représentatives de ce transcrit :

La première séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°48.

20 La séquence SEQ ID N°48 a une longueur de 373 nucléotides.

La seconde séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°49.

25 La séquence SEQ ID N°49 a une longueur de 459 nucléotides.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

30 La troisième séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N° 50.

La séquence SEQ ID N° 50 a une longueur de 2575 nucléotides

Dès homologues entre la séquence SEQ ID N° 50 et des séquences référencées dans la base de données GenBank (Version 116) ont été observées. Ces homologues sont les suivantes :

- 5 - 98% d'identité sur 720 pb avec un BAC END g6348761 AQ892571 HS_3143_A1_G01_T7C CIT Approved Human Genomic Sperm Library length 848pb

10 Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

15 **Gène GS929341.**

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS929341. Deux séquences nucléiques représentatives de ce transcrit ont été déterminées.

20 La première séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°51.

La séquence SEQ ID N°51 a une longueur de 231 nucléotides.

25 La seconde séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°52.

La séquence SEQ ID N°52 a une longueur de 344 nucléotides. Cette séquence comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 3 au nucléotide en position 131 de la séquence SEQ ID N°52.

Aucune identité avec les séquences SEQ ID N° 51 et 52 n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

5 La troisième séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N° 53.

 La séquence SEQ ID N° 53 a une longueur de 402 nucléotides. Cette séquence comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 1 au nucléotide en position 188 de la séquence
10 SEQ ID N° 53.

 Aucune identité avec les séquences SEQ ID N° 51 à 53 n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 116).

15 Les analyses d'expression du transcrit des séquences SEQ ID N°51 et 52 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+de différents tissus ont permis de montrer que le gène GS929341 était exprimé dans le cerveau fœtal, le foie, le cerveau, le cœur, la prostate, le placenta, le foie fœtal,
20 l'utérus, le testicule, le rein, le muscle squelettique et le poumon.

 Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences
25 familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Polypeptide codé par l'acide nucléique de séquence SEQ ID NO 52

 Le cadre ouvert de lecture de la séquence d'acide nucléique SEQ
30 ID N°52 code potentiellement pour un polypeptide de 43 aminoacides de longueur constituant la séquence SEQ ID N°156.

Aucune homologie significative avec BLAST n'a été retrouvée avec les séquences référencées dans les bases de données Swissprot (version 36, dernière remise à jour du 3 mai 1999) et PRODOM: (domaines homologues détectés dans Swissprot, version 34.2, 5 novembre 1997).

Polypeptide codé par l'acide nucléique de séquence SEQ ID N° 53

Le cadre ouvert de lecture de la séquence d'acide nucléique SEQ ID N° 53 code potentiellement pour un polypeptide de 61 acides aminés 10 de longueur constituant la séquence SEQ ID N° 157.

Aucune homologie significative avec BLAST n'a été retrouvée avec les séquences référencées dans les bases de données Swissprot (version 38), PIR (Version 62, Septembre 1999), trEMBL (Version de Août 1999) et Gen Pept (Version 115).

15 Les polypeptides de séquences SEQ ID N°156 et 157 sont susceptibles d'intervenir dans la régulation du flux du cholestérol, et plus particulièrement de la maladie de Tangier, dans des déficiences familiales en HDL, ou encore dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

20

Gène GS915742.

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messager correspondant à 25 un transcrit du gène désigné ici sous le n° GS915742. Trois séquences nucléiques représentatives de ce transcrit ont été déterminées.

La première séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°54.

La séquence SEQ ID N°54 a une longueur de 228 nucléotides.

30

La seconde séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°55.

La séquence SEQ ID N°55 a une longueur de 270 nucléotides.

5 La troisième séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N° 56.

La séquence SEQ ID N° 56 a une longueur de 1130 nucléotides.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Versions 110 et 116).

10 Les analyses d'expression du transcrit des séquences SEQ ID N°54 et 55 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS915742 était exprimé dans le cerveau
15 fœtal, le foie, le placenta et le rein.

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus
20 particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

25 **Gène GS913018.**

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS913018. Deux séquences représentatives de ce transcrit sont représentées ci-après.

30 La première séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°57.

La séquence SEQ ID N°57 a une longueur de 463 nucléotides.

La seconde séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°58.

5 La séquence SEQ ID N°58 a une longueur de 289 nucléotides.

Aucune identité de séquence avec les séquences SEQ ID N° 57 et 58 n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

10 A partir de la séquence SEQ ID N° 57, on a synthétisé deux amorces nucléotidiques, respectivement les amorces de séquences SEQ ID N° 111 et 112.

A partir de la séquence SEQ ID N° 58, on a synthétisé deux amorces nucléotidiques, respectivement les amorces de séquences SEQ ID N° 113 et 114.

15 Les amorces de séquences SEQ ID N° 111 à 114 ont permis d'amplifier un ADNc, à partir d'une banque d'ARNm polyA+ de différents tissus humains commercialisée par la Société Clontech une troisième séquence nucléique de l'ADNc correspondant au transcrit du gène GS913018.

20

La troisième séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N° 59.

La séquence SEQ ID N° 59 a une longueur de 1542 nucléotides.

25 Des homologues ont été retrouvées sur la séquence SEQ ID N° 59, en particulier différents fragments d'identité sur les positions [735-1268] [1-357] [559-710] et [373-501] de la séquence SEQ ID N° 59 avec la séquence g6563616 (AC013740)Homo sapiens clone RP11-115J22, WORKING DRAFT SEQUENCE, 15 unordered pieces.Length = 180711, réépertoriées dans GenBank (Version 116).

Les analyses d'expression du transcrit des séquences SEQ ID N°57 et 58 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS913018 était exprimé dans le cerveau
5 fœtal, le foie, le cerveau, le cœur, la prostate, le placenta, le foie fœtal, l'utérus, le testicule et le rein.

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus
10 particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

15 **Gène GS911742.**

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS911742. Trois séquence représentatives de ce transcrit ont été déterminées.

20 La première séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°60.

La séquence SEQ ID N°60 a une longueur de 1417 nucléotides.

La seconde séquence nucléique de l'ADNc correspondant à ce
25 transcrit constitue la séquence SEQ ID N°61.

La séquence SEQ ID N°61 a une longueur de 696 nucléotides.

Aucune identité de séquence avec les séquences SEQ ID N° 60 et 61 n'a été retrouvée lors d'une recherche dans la base de données
30 GenBank (Version 110).

La troisième séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N° 62.

La séquence SEQ ID N° 62 a une longueur de 2702 nucléotides. Cette séquence comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 1 au nucléotide en position 792 de la séquence SEQ ID N° 62. La séquence codante potentielle débute au nucléotide en position 49 et se termine au nucléotide en position 792 de la séquence SEQ ID N° 62. Le codon d'initiation de la traduction débute au nucléotide en position 49 de la séquence SEQ ID N° 62. Cette séquence comprend un motif de Kozak de séquence " GC CGC GCC ATG C " qui débute au nucléotide en position 41 de la séquence SEQ ID N° 62.

Des homologies avec la séquence SEQ ID N° 62 ont été observées avec des séquences répertoriées dans la base de données GenBank (Version 116). Ces homologies sont les suivantes :

- 15 • 98 % d'homologie sur 1410 pdb (positions [4-1413]) avec la séquence gi|5912095|embl|AL117557.1|HSM801083[5912095] Homo sapiens mRNA; cDNA DKFZp564D177 (from clone DKFZp564D177); partial cds. Length = 1431; deposee le 18-FEB-2000 ; Submitted (15-SEP-1999) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY, Bloecker,H., Boecher,M., Brandt,P., Wiemann,S.
- 25 • 97 % d'homologie sur 10139 pdb [1-1039] et 97% sur 380 pdb [1082-1458] et 90% sur 51 pdb [1506-1556] avec la séquence gi|6841247|gb|AF161417.1|AF161417[6841247] Homo sapiens HSPC299 mRNA, partial cds. Length = 1659; deposee le 01-FEB-2000 ; Direct Submission ; Submitted (14-MAY-1999) Shanghai Institute of Hematology, Shanghai Second Medical University, Rui-Jin Hospital, 197 Rui-Jin Road II,
- 30

- 98 % d'homologie sur 1410 pdb [4-1413] avec la sequence g5912095 (AL117557) Homo sapiens mRNA; cDNA DKFZp564D177 (from clone DKFZp564D177); partial cds. Length = 1431; deposee le 18-FEB-2000 ; Direct Submission ; Submitted (15-SEP-1999) MIPS, Am Klopferspitz 18a, D-82152 ; Martinsried, GERMANY, Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
- 93% sur 911 pdb (positions [1-911]) et 93% sur 179 pdb(positions [1395-1573]) et 81% sur 131 pdb (positions [992-1122]) avec la sequence gi|7023832|dbj|AK002137.1|AK002137[7023832] Homo sapiens cDNA FLJ11275 fis, clone PLACE1009375. Length = 1564; deposee le 22-FEB-2000 ; NEDO human cDNA sequencing project ; Unpublished
- Differents fragments d'homologie(de 90% a 100%) avec la sequence g5932616 (AC009594) Homo sapiens chromosome 4 clone 363_G_01 map 4, *** SEQUENCING IN PROGRESS ***, 9 unordered pieces. Length = 150108;

20

Les analyses d'expression du transcrit des séquences SEQ ID N°60 et 61 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+de différents tissus ont permis de montrer que le gène GS911742 était exprimé dans le cerveau foetal, le foie, le cœur et le placenta.

25

De plus, une analyse de l'expression du transcrit par Northern blot, , selon le protocole décrit de l'Exemple 1, à l'aide de la sonde de séquence SEQ ID NO 101, a révélé la présence de transcrits dans le blot commercialisé par la Société Clontech (Ref. N° 7759-1).

30

La taille des transcrits détectés avec la sonde de séquence SEQ ID N°101 est de 1,9 kb dans le pancréas, le rein, le muscle squelettique, le poumon et le placenta.

- 5 Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL.

10 ***Polypeptide codé par l'acide nucléique de séquence SEQ ID NO 62***

Le cadre ouvert de lecture partiel de la séquence d'acide nucléique SEQ ID N°62 code potentiellement pour un polypeptide de 263 aminoacides de longueur constituant la séquence SEQ ID N°158.

- 15 Homologies observées au niveau de la séquence protéique traduite en six phases :

- 99% d'homologie sur 262 AA (position [4-789]) avec la sequence g5912096 (AL117557) hypothetical protein [Homo sapiens] Length = 263 ; note= "similarity to NIPSNAP1" ; deposee le 18-FEB-2000 ; Direct Submission ; Submitted (15-SEP-1999) MIPS, Am Klopferspitz
20 18a, D-82152 , Martinsried, GERMANY, Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and Wiemann,S,
- Homologies avec des sequences NISNAP2 et NISNAP2 :
27 % d'homologie sur 179 AA avec la sequence g2769254
25 (AJ001259) NIPSNAP2 protein [Homo sapiens] Length = 285 et 24% d'homologie sur 211 AA avec la sequence g2769649 (AJ001258) NIPSNAP1 protein [Homo sapiens] Length = 284
- 27 % d'homologie sur 179 AA avec la sequence g3403167
30 (AF029786) GBAS [Homo sapiens]

Length = 286 . GBAS : guanine nucleotide-binding protein, alpha-subunit (adenylate cyclase-stimulating G alpha protein).

Cette proteine est decrite avec un site de phosphorylation et une region transmembranaire.

- 5 Les " guanine nucleotide-binding proteins " ont un role de modulateurs ou transducteurs dans differents sytemes de signalisation transmembranaires.

- 21% d'identite avec la proteine YMQ1_CAEEL (Base de donnees
10 Prodom, VERSION ?) qui a une simiralite avec la proteien SNAP25 et la 4-nitrophenylphosphatase.

- Le polypeptide de sequence SEQ ID N° 158 comprend
des sites de glycosylation, des sites de phosphorylation tels que des
15 sites de phosphorylation cAMP- et cGMP-dépendants de proteine kinase et de caséine kinese de type II.

- 20 Fonction putative :

- Le polypeptide de sequence SEQ ID N°158 s'apparente donc par son motif retrouvé dans le domaine PD013981 de PRODOM: (domaines homologues détectés dans Swissprot, version 34.2, novembre 1997) à une proteine putativement impliquée dans le transport vésiculaire
25 intracellulaire. Ce mecanisme sous-tendu par un adressage spécifique médié par ces proteines est d'intérêt par rapport à la maladie de Tangier/FHD qui est décrite par un défaut du mecanisme de transport et de translocation des pools de cholesterol intracellulaire vers les particules HDL acceptrices. Le polypeptide de sequence SEQ ID N°158

est donc susceptible d'intervenir dans une étape importante impliquée dans le transport inverse du cholestérol par les HDL.

Le polypeptide de séquence SEQ ID N° 158 est susceptible
5 d'intervenir dans la régulation du flux du cholestérol, et plus particulièrement de la maladie de Tangier, dans des déficiences familiales en HDL, ou encore dans une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

10 **Gène GS98601.**

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messenger correspondant à un transcrit du gène désigné ici sous le n° GS98601. Trois séquences nucléiques représentatives de ce transcrit ont été déterminées.

15 La première séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°63.

La séquence SEQ ID N°63 a une longueur de 335 nucléotides.

Aucune identité de séquence n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

20

La seconde séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°64.

La séquence SEQ ID N°64 a une longueur de 447 nucléotides.

Aucune identité de séquence n'a été retrouvée lors d'une
25 recherche dans la base de données GenBank (Version 110).

La troisième séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°65.

La séquence SEQ ID N°65 a une longueur de 2324 nucléotides.

Cette séquence comprend un cadre ouvert de lecture (ORF) partiel allant du nucléotide en position 3 au nucléotide en position 611 de la séquence SEQ ID N°65.

Homologies observées au niveau de la séquence nucléotidique :

5 (307)

- 99% d'homologie sur 514 pb (position 1508-2021 pb) avec des séquences de GenBank: gi|3483520| correspondant au clone ZB95F02 d'ADNc (Homo sapiens) dont la séquence est complète.
- 98% d'homologie sur 170 pb (position 862-1031 pb) avec des séquences de GenBank: gi|1184671| (région 3'UTR partielle de l'ARN m, codant pour la protéine inductible du virus de la maladie de newcastle)

Les analyses d'expression du transcrit de séquence SEQ ID N°45 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS98601 était exprimé dans le cerveau, le placenta et l'utérus.

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Polypeptide codé par l'acide nucléique de séquence SEQ ID NO 65

Le cadre ouvert de lecture de la troisième séquence d'acides nucléiques SEQ ID N°65 code potentiellement pour un polypeptide de 203 aminoacides de longueur constituant la séquence SEQ ID N°159.

- 5 Homologies observées au niveau de la séquence protéique (307):
34% d'homologie sur 180 aa (position 3-182 aa) avec des séquences de
Genpept: (traduction de Genbank v110 et 111, dernière remise à jour le
7 mai 1999): gi|3878571|gnl|PID|e1348103 (Z46381)- faible similarité
avec la protéine de levure Ysy6 - (PIR: (SEQUENCES PIR NON
10 REDONDANTES, VERSION 57) numéro d'accès JQ0912); EST d'ADNc
EMBL:D32318 provient de ce gène ; EST d'ADNc EMBL:D33688
provient de ce gène EST d'ADNc EMBL:D34664 provient de ce gène ;
EST d'ADNc EMBL:D36574 provient de ce gène et des séquences de
SP-TrEMBL (SP-TrEMBL, version 7, novembre 1998): sp|Q21453|
15 correspondant à la protéine M01F1.4 de C elegans

- Le polypeptide de séquence SEQ ID N° 159 est susceptible
d'intervenir dans la régulation du flux du cholestérol, et plus
particulièrement de la maladie de Tangier, dans des déficiences
20 familiales en HDL, ou encore dans une maladie génétiquement liée au
locus 9q31-34 du chromosome 9.

Fonction putative :

- Ce gène est un candidat par sa localisation chromosomique pour les
25 pathologies Tangier/FHD.

Gène GS94852.***Acide(s) nucléique(s)***

Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS94852. Trois séquences nucléiques représentatives de ce transcrit ont été déterminées.

La première séquence nucléique de l'ADNc correspondant à ce
5 transcrit constitue la séquence SEQ ID N°66.

La séquence SEQ ID N°66 a une longueur de 447 nucléotides.

La seconde séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°67.

10 La séquence SEQ ID N°67 a une longueur de 564 nucléotides.

A partir de la séquence SEQ ID N° 66, on a synthétisé deux amorces nucléotidiques, respectivement les amorces de séquences SEQ ID N° 115 et 116.

15 A partir de la séquence SEQ ID N° 67, on a synthétisé deux amorces nucléotidiques, respectivement les amorces de séquences SEQ ID N° 117 et 118

Les amorces de séquences SEQ ID N° 115 à 118 ont permis d'amplifier un ADNc, à partir d'une banque d'ARNm polyA+ de différents
20 tissus humains commercialisée par la Société Clontech une troisième séquence nucléique de l'ADNc correspondant au transcrit du gène GS94852.

La troisième séquence nucléique correspondant au transcrit du gène GS94852 constitue la séquence SEQ ID N° 68.

25 La séquence SEQ ID N° 68 a une longueur de 604 nucléotides.

Aucune identité de séquence avec les séquences SEQ ID N° 66 à 68 n'a été retrouvée lors d'une recherche dans la base de données GenBank (Versions 110 et 116).

Les analyses d'expression du transcrit de séquence SEQ ID N°67 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+ de différents tissus ont permis de montrer que le gène GS94852 était exprimé dans le foie et le coeur.

5

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

10

Gène S935135.

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS935135. Trois séquences nucléiques représentatives de ce transcrit ont été déterminées.

15

La première séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°69.

La séquence SEQ ID N°69 a une longueur de 482 nucléotides.

20

La seconde séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°70.

La séquence SEQ ID N°70 a une longueur de 402 nucléotides.

Aucune identité de séquence n'a été retrouvée avec les séquences SEQ ID N° 69 et 70 lors d'une recherche dans la base de données GenBank (Version 110).

25

On a synthétisé une première amorce nucléotidique de séquence SEQ ID N° 119 à partir de la séquence SEQ ID N°69 et une seconde amorce nucléotidique de séquence SEQ ID N° 120 à partir de la

30

séquence SEQ ID N° 70. Ces amorces ont permis d'amplifier une troisième séquence nucléique représentative du transcrit du gène GS935135 constituant la séquence SEQ ID N° 71.

La séquence nucléique SEQ ID N° 71 a une longueur de 758 nucléotides.

Des homologies ont été retrouvées avec des séquences répertoriées dans la base de données GenBank (Version 116). Ces homologies sont les suivantes :

- 80 a 85% d'homologie sur 3 fragments (156+197+93 pdb) avec la séquence g2168141 (gi|2168141|emb|Z93019.1|HS49C23[2168141) Human DNA sequence from PAC 49C23 on chromosome X contains malate dehydrogenase pseudogene and STS. Length = 153078

15

- 81% a 90% d'homologie sur 4 fragments (144+86+197+137 pdb) avec la séquence g2828782 (gi|2828782|gb|AC002319.1|AC002319[2828782]) Homo sapiens chromosome 9q34, clone 70C11, complete sequence. Length = 46305

20

Les analyses d'expression du transcrit de séquence SEQ ID N°69 ou 70 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+de différents tissus ont permis de montrer que le gène GS935135 était exprimé dans le cerveau foetal, le foie, le cerveau, la prostate, le placenta, le foie foetal, l'utérus, le testicule et le rein.

30

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus
5 9q31-34 du chromosome 9.

Gène GS914669.

Acide(s) nucléique(s)

10 Il a été isolé selon l'invention un ARN messenger correspondant à un transcrit du gène désigné ici sous le n° GS914669. Trois séquences nucléiques représentatives de ce transcrit ont été déterminées.

La première séquence nucléique de l'ADNc correspondant à ce
15 transcrit constitue la séquence SEQ ID N°72.

La séquence SEQ ID N°72 a une longueur de 673 nucléotides.

La seconde séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°73.

20 La séquence SEQ ID N°73 a une longueur de 554 nucléotides.

Aucune identité de séquence avec les séquences SEQ ID N° 72 et 73 n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

25 A partir de la séquence SEQ ID N° 72, on a synthétisé deux amorces nucléotidiques, respectivement les amorces de séquences SEQ ID N° 121 et 122.

A partir de la séquence SEQ ID N° 73, on a synthétisé deux amorces nucléotidiques, respectivement les amorces de séquences
30 SEQ ID N° 123 et 124.

.Les amorces de séquences SEQ ID N° 121 à 124 ont permis d'amplifier un ADNc, à partir d'une banque d'ARNm polyA+ de différents tissus humains commercialisée par la Société Clontech une troisième séquence nucléique de l'ADNc correspondant au transcrit du gène GS914669. Cette séquence constitue la séquence SEQ ID N° 74. La
5 séquence SEQ ID N° 74 a une longueur de 1794 nucléotides. Elle comprend un cadre ouvert de lecture allant du nucléotide en position 1 au nucléotide en position 258 de la séquence SEQ ID N° 74 ainsi qu'une séquence codante localisée identiquement. Cette séquence comprend
10 un site de polyadénylation débutant au nucléotide en position 1751 de la SEQ ID N° 74.

Des homologies avec la séquence SEQ ID N° 74 ont été retrouvées dans les séquences répertoriées dans la base de données GenBank (Version 116). Ces homologies sont les suivantes :
15 99% d'identité sur 1000pb (792 a 1793pb) g6807977 AL137422 Homo sapiens mRNA; cDNA DKFZp761A1623 (from clone DKFZp761A1623); partial cds Length = 1000

Identité avec un BAC en cours de séquençage AL137023 g6982086
20 Homo sapiens chromosome 9 clone RP11-403A22 map q34.13-34.3, *** SEQUENCING IN PROGRESS ***, 19 unordered pieces. Length = 184814

Les analyses d'expression du transcrit de séquence SEQ ID N°72
25 ou 73 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+de différents tissus ont permis de montrer que le gène GS914669 était exprimé dans le cerveau foetal et le coeur.

30 Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences

familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Polypeptide codé par l'acide nucléique de séquence SEQ ID N° 74

5 Le cadre de lecture ouvert de la séquence d'acide nucléique SEQ ID N° 74 code potentiellement pour le polypeptide de séquence SEQ ID N° 160 d'une longueur de 85 acides aminés.

Aucune homologie de séquence n'a été observée avec les séquences répertoriées dans les bases de données Genpept (Version 10 115), Swissprot (Version 38), trEMBL (Version de Août 1999) et PIR (Version 62 de Septembre 1999).

Gène GS913839.

Acide(s) nucléique(s)

15 Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS913839. Trois séquences nucléiques représentatives de ce transcrit ont été déterminées.

La première séquence nucléique de l'ADNc correspondant à ce
20 transcrit constitue la séquence SEQ ID N°75.

La séquence SEQ ID N°75 a une longueur de 507 nucléotides.

La seconde séquence nucléique de l'ADNc correspondant à ce
transcrit constitue la séquence SEQ ID N°76.

25 La séquence SEQ ID N°76 a une longueur de 415 nucléotides.

Aucune identité de séquence avec les séquences SEQ ID N° 75 et 76 n'a été retrouvée lors d'une recherche dans la base de données GenBank (Version 110).

A partir de la séquence SEQ ID N° 75, on a synthétisé une amorce nucléotidique de séquence SEQ ID N° 125.

A partir de la séquence SEQ ID N° 76, on a synthétisé une amorce nucléotidique de séquence SEQ ID N° 1126

5 Les amorces de séquences SEQ ID N° 125 et 126 ont permis d'amplifier un ADNc, à partir d'une banque d'ARNm polyA+ de différents tissus humains commercialisée par la Société Clontech une troisième séquence nucléique de l'ADNc correspondant au transcrit du gène GS94852. Cette séquence constitue la séquence SEQ ID N° 77.

10 La séquence SEQ ID N° 77 a une longueur de 1318 nucléotides.

Des homologies de la séquence SEQ ID N° 77 ont été observées avec des séquences répertoriées dans la base de données GenBank (Version 116). Ces homologies sont les suivantes :

15

- 99% d'homologie sur 1320 pdb (positions [1-1318]) avec la sequence g6006243 (AC011096) Homo sapiens clone 2_D_21, *** SEQUENCING IN PROGRESS ***, 15 unordered pieces.Length = 135130;

20

- 99% d'homologie sur 1320 pdb (positions [1-1318]) avec la sequence g7263520 (AL161631) Homo sapiens chromosome 9 clone RP11-70K10, *** SEQUENCING IN PROGRESS ***, 45 unordered pieces.

25

Length = 100562;

Les analyses d'expression du transcrit de séquence SEQ ID N°75 ou 76 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1.

30 Ces analyses effectuées à partir d'ARN polyA+de différents tissus ont

permis de montrer que le gène GS913839 était exprimé dans le cerveau fœtal et le foie.

5 Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

10

Gène GS912639.

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messenger correspondant à un transcrit du gène désigné ici sous le n° GS912639. Trois séquences
15 nucléiques représentatives de ce transcrit ont été déterminées.

La première séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°78.

La séquence SEQ ID N°78 a une longueur de 530 nucléotides.

20 La seconde séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°79.

La séquence SEQ ID N°79 a une longueur de 495 nucléotides.

Aucune identité de séquence avec les séquences SEQ ID N° 78 et 79 n'a été retrouvée lors d'une recherche dans la base de données
25 GenBank (Version 110).

A partir de la séquence SEQ ID N° 78, on a synthétisé une amorce nucléotidique de séquence SEQ ID N° 127.

A partir de la séquence SEQ ID N° 79, on a synthétisé une amorce nucléotidique de séquence SEQ ID N° 128

Les amorces de séquences SEQ ID N° 127 et 128 ont permis d'amplifier un ADNc, à partir d'une banque d'ARNm polyA+ de différents tissus humains commercialisée par la Société Clontech une troisième séquence nucléique de l'ADNc correspondant au transcrit du gène GS912639. Cette séquence constitue la séquence SEQ ID N° 80.

La séquence SEQ ID N° 80 a une longueur de 594 nucléotides.

Des homologies de séquence de la sEQ ID N° 80 ont été retrouvées avec des séquences référencées dans la base de données GenBank (Version 116). Ces homologies sont les suivantes :

- 99% d'homologie sur 522 pdb (positions [204-725]) avec la séquence g2603415 (gi|2603415|gb|B51178.1|B51178[2603415]) CIT978SK-95K15.TV CIT978SK Homo sapiens genomic clone 95K15, genomic survey sequence.Length = 524;
- 99% d'homologie sur 501 pdb(positions [204-704]) avec la séquence g2866378 (gi|2866378|gb|B79355.1|B79355[2866378]) CIT978SK-95K15.TV.1 CIT978SK Homo sapiens genomic clone 95K15, genomic survey sequence.Length = 529;
- 94% d'homologie sur 309 pdb (positions [205-513]) avec la séquence g2602442 (i|2602442|gb|B50205.1|B50205[2602442]) CIT978SK-96F5.TV CIT978SK Homo sapiens genomic clone 96F5,genomic survey sequence.Length = 309;

Les analyses d'expression du transcrit de séquence SEQ ID N°78 ou 79 ont été réalisées par RT PCR, comme décrit dans l'Exemple 1. Ces analyses effectuées à partir d'ARN polyA+de différents tissus ont permis de montrer que le gène GS912639 était exprimé dans le foie.

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Gène GS933630.

Acide(s) nucléique(s)

Il a été isolé selon l'invention un ARN messager correspondant à un transcrit du gène désigné ici sous le n° GS933630. Une séquence nucléique représentative de ce transcrit a été déterminée.

Cette séquence nucléique de l'ADNc correspondant à ce transcrit constitue la séquence SEQ ID N°81.

La séquence SEQ ID N° 81 a une longueur de 582 nucléotides.

Aucune homologie n'a été observée avec les séquences référencées dans la base de données GenBank (Version 116).

Ce gène constitue un candidat positionnel causal d'une maladie due à un dysfonctionnement du flux inverse du cholestérol, et plus particulièrement de la maladie de Tangier ou encore des déficiences familiales en HDL, ou encore à une maladie génétiquement liée au locus 9q31-34 du chromosome 9.

Caractéristiques de l'invention

L'invention concerne ainsi un acide nucléique codant pour une protéine ayant une séquence en acides aminés choisie dans le groupe des séquences d'acides aminés SEQ ID N°129 à SEQ ID N°160 ou un fragment peptidique ou un variant de cette dernière ou un acide nucléique de séquence complémentaire.

De manière générale, les acides nucléiques selon l'invention se présentent sous une forme isolée ou purifiée.

5 L'invention concerne également un acide nucléique comprenant au moins huit nucléotides consécutifs d'un polynucléotide choisi dans le groupe constitué des séquences nucléotidiques SEQ ID N°1 à SEQ ID N°81 et SEQ ID N°82 à SEQ ID N°101, ou un acide nucléique de séquence complémentaire.

10 L'invention est aussi relative à un acide nucléique comprenant au moins 20, 30, 40, 50, 100 ou 150 nucléotides consécutifs d'un polynucléotide choisi dans le groupe constitué des séquences nucléotidiques SEQ ID N°1 à SEQ ID N°81 et SEQ ID N°82 à SEQ ID N°101, ou un acide nucléique de séquence complémentaire.

15 Selon un autre aspect, l'invention concerne aussi un acide nucléique ayant au moins 90% d'identité en nucléotides avec un acide nucléique choisi dans le groupe constitué des séquences nucléotidiques SEQ ID N°1 à SEQ ID N°81 et SEQ ID N°82 à SEQ ID N°101, avantageusement 80%, de préférence 95, 99%, 99,5%, et de manière tout à fait préférée 99,8% d'identité en nucléotides avec un acide
20 nucléique choisi dans le groupe constitué des séquences nucléotidiques SEQ ID N°1 à SEQ ID N°81 et SEQ ID N°82 à SEQ ID N°101, ou un acide nucléique de séquence complémentaire.

Selon encore un autre aspect, l'invention est relative à un acide nucléique hybridant, dans des conditions d'hybridation de forte
25 stringence, avec un acide nucléique tel que défini ci-avant, et plus particulièrement un acide nucléique choisi dans le groupe constitué des séquences nucléotidiques SEQ ID N°1 à SEQ ID N°81 et SEQ ID N°82 à SEQ ID N°101, ou un acide nucléique de séquence complémentaire.

30 Comme cela a été décrit en détails plus haut, chacune des séquences nucléotidiques SEQ ID N°1 à SEQ ID N°81 constitue des ADNc dont la séquence nucléotidique est retrouvée dans les transcrits de gènes potentiellement impliqués dans des affections liées à un dysfonctionnement du métabolisme des lipoprotéines, en particulier le
35 transport inverse du cholestérol.

Pour certains de ces acides nucléiques, une phase de lecture ouverte a été déterminée, qui code pour un polypeptide dont une altération dans la séquence d'acides aminés ou dans l'expression est potentiellement associée à l'une de ces affections, ce qui indique que les

5 séquences nucléotidiques comprenant les phases de lecture ouvertes constituent des acides nucléiques d'intérêt potentiellement thérapeutique.

En conséquence, l'invention a en outre pour objet un acide nucléique ayant au moins 80% d'identité en nucléotides avec un

10 polynucléotide comprenant, ou alternativement constitué de, un cadre de lecture ouvert complet ou partiel, tel que défini ci-avant dans la présente description.

Les acides nucléiques ci-dessus qui comprennent, totalement ou partiellement, la région codante des produits de transcription de

15 séquences SEQ ID N°1 à SEQ ID N°81 peuvent être exprimés dans une cellule hôte désirée, lorsque ces acides nucléiques sont placés sous le contrôle de signaux d'expression convenables.

De tels signaux d'expression peuvent être indifféremment les signaux d'expression contenus dans les régions régulatrices de chacun

20 des gènes correspondants ou au contraire constitués des séquences nucléiques régulatrices exogènes.

Un tel acide nucléique placé sous le contrôle d'une séquence régulatrice fonctionnelle dans la cellule hôte désirée peut être aussi inséré dans un vecteur en vue de son expression.

25

SONDES ET AMORCES NUCLEOTIDIQUES

Les fragments d'acides nucléiques dérivés de l'une quelconque des séquences nucléotidiques SEQ ID N°1 à SEQ ID N°81 sont utiles

30 pour la détection de la présence d'au moins une copie d'une séquence nucléotidique choisie parmi les séquences SEQ ID N°1 à SEQ ID N°81 ou encore d'un fragment ou d'un variant de cette dernière dans un échantillon.

Les sondes ou les amorces nucléotidiques selon l'invention

35 comprennent au moins huit nucléotides consécutifs d'un acide nucléique

choisi dans le groupe constitué des séquences SEQ ID N°1 à SEQ ID N°81, ou d'un acide nucléique de séquence complémentaire.

De préférence, des sondes ou amorces nucléotidiques selon l'invention auront une longueur de 10, 12, 15, 18 ou 20 à 25, 35, 40, 50, 70, 80, 100, 200, 500, 1000, 1500 nucléotides consécutifs d'un acide nucléique selon l'invention, en particulier un acide nucléique de séquence nucléotidique choisie parmi les séquences SEQ ID N°1 à SEQ ID N°81 ou d'un acide nucléique de séquence complémentaire.

Alternativement, une sonde ou une amorce nucléotidique selon l'invention consistera et/ou comprendra les fragments d'une longueur de 12, 15, 18, 20, 25, 35, 40, 50, 100, 200, 500, 1000, 1500 nucléotides consécutifs d'un acide nucléique selon l'invention, plus particulièrement d'un acide nucléique choisi parmi les séquences SEQ ID N°1 à SEQ ID N°81, ou d'un acide nucléique de séquence complémentaire.

La définition d'une sonde et d'une amorce nucléotidique selon l'invention englobe donc des oligonucléotides qui hybrident, dans les conditions d'hybridation de forte stringence définies ci-avant, avec un acide nucléique choisi parmi les séquences SEQ ID N°1 à SEQ ID N°81 ou avec une séquence complémentaire de ces derniers.

Des sondes et amorces préférées selon l'invention comprennent tout ou partie d'un polynucléotide choisi parmi les séquences nucléotidiques SEQ ID N°82 à 101, ou des acides nucléiques de séquence complémentaire, ou encore parmi les séquences nucléotidiques SEQ ID N° 102 à 128, ou des acides nucléiques de séquence complémentaire.

Une amorce ou une sonde nucléotidique selon l'invention peut être préparée par toute méthode adaptée bien connue de l'homme du métier, y compris par clonage et action d'enzymes de restriction ou encore par synthèse chimique directe selon des techniques telles que la méthode au phosphodiester de **NARANG et al. (1979)** ou de **BROWN et al. (1979)**, la méthode aux diéthylphosphoramidites de **BEAUCAGE et al. (1980)** ou encore la technique sur support solide décrite dans le brevet EU N°EP 0 707 592.

Chacun des acides nucléiques selon l'invention, y compris les sondes et amorces oligonucléotidiques décrites ci-dessus, peuvent être

marqués, si désiré, en incorporant un marqueur détectable par des moyens spectroscopiques, photochimiques, biochimiques, immunochimiques ou encore chimiques.

Par exemple, de tels marqueurs peuvent consister en des isotopes radioactifs (^{32}P , ^{33}P , ^3H , ^{35}S), des molécules fluorescentes (5-bromodeoxyuridine, fluorescéine, acétylaminofluorène, digoxigénine) ou encore des ligands tels que la biotine.

Le marquage des sondes est fait de préférence par incorporation de molécules marquées au sein des polynucléotides par extension d'amorces, ou bien par rajout sur les extrémités 5' ou 3'

Des exemples de marquage non radioactifs de fragments d'acides nucléiques sont décrits notamment dans le brevet français n° FR 78 109 75 ou encore dans les articles de URDEA et al. (1988) ou SANCHEZ-PESCADOR et al. (1988).

De manière avantageuse, les sondes selon l'invention peuvent avoir des caractéristiques structurales de nature à permettre une amplification du signal, telles que les sondes décrites par URDEA et al. (1991) ou encore dans le brevet européen n° EP-0 225 807 (CHIRON).

Les sondes oligonucléotides selon l'invention peuvent être utilisées notamment dans des hybridations de type Southern à l'ADN génomique ou encore dans des hybridations à l'ARN messager correspondant lorsque l'expression du transcrit correspondant est recherchée dans un échantillon.

Les sondes selon l'invention peuvent aussi être utilisées pour la détection de produits d'amplification PCR ou encore pour la détection de mésappariements.

Des sondes ou amorces nucléotidiques selon l'invention peuvent être immobilisées sur un support solide. De tels supports solides sont bien connus de l'homme du métier et comprennent des surfaces des puits de plaques de microtitration, des lits de polystyrène, des lits magnétiques, des bandes de nitrocellulose, ou encore des microparticules telles que des particules de latex.

En conséquence, la présente invention concerne également un procédé de détection de la présence d'un acide nucléique tel que décrit ci-avant dans un échantillon, ladite méthode comprenant les étapes de :

1) mettre en contact une ou plusieurs sondes nucléotidiques selon l'invention avec l'échantillon à tester ;

2) détecter le complexe éventuellement formé entre la ou les sondes et l'acide nucléique présent dans l'échantillon.

5

Selon un mode de réalisation particulier du procédé de détection selon l'invention, la ou les sondes oligonucléotidiques sont immobilisées sur un support.

10 Selon un autre aspect, les sondes oligonucléotidiques comprennent un marqueur détectable.

L'invention concerne en outre un nécessaire ou kit pour la détection de la présence d'un acide nucléique selon l'invention dans un échantillon, ledit nécessaire comprenant :

15 a) une ou plusieurs sondes nucléotidiques telles que décrites ci-dessus ;

b) le cas échéant, les réactifs nécessaires à la réaction d'hybridation.

20 Selon un premier aspect, le nécessaire ou kit de détection est caractérisé en ce que la ou les sondes sont immobilisées sur un support.

Selon un second aspect, le nécessaire ou kit de détection est caractérisé en ce que les sondes oligonucléotidiques comprennent un marqueur détectable.

25 Selon un mode de réalisation particulier du kit de détection décrit ci-dessus, un tel kit comprendra une pluralité de sondes oligonucléotidiques conformes à l'invention qui pourront être utilisées pour détecter des séquences cibles d'intérêt ou alternativement détecter des mutations dans les régions codantes ou les régions non codantes des acides nucléiques selon l'invention, plus particulièrement des acides nucléiques de séquences SEQ ID N°1 à SEQ ID N°81 ou les acides nucléiques de séquence complémentaire.

30

Des sondes préférées comprendraient tout ou partie des polynucléotides de séquences SEQ ID N°82 à SEQ ID N°101.

35 Ainsi, les sondes selon l'invention immobilisées sur un support peuvent être ordonnées en matrices telles que les " puces à ADN ". De

telles matrices ordonnées ont été en particulier décrites dans le brevet US N° 5,143,854, dans les demandes PCT N° WO 90/150 70 et 92/10092.

Des matrices supports sur lesquelles des sondes
5 oligonucléotidiques ont été immobilisées à une haute densité sont par exemple décrites dans les brevets US N°5,412,087 et dans la demande PCT N°WO 95/11995.

Les amorces nucléotidiques selon l'invention peuvent être utilisées pour amplifier l'un quelconque des acides nucléiques selon
10 l'invention, et plus particulièrement tout ou partie d'un acide nucléique de séquences SEQ ID N°1 à SEQ ID N°81, ou encore un variant de celui-ci.

Un autre objet de l'invention concerne un procédé pour l'amplification d'un acide nucléique selon l'invention, et plus particulièrement un acide nucléique de séquences SEQ ID N°1 à SEQ ID
15 N°81 ou un fragment ou un variant de celui-ci contenu dans unchantillon, ledit procédé comprenant les étapes de :

a) mettre en contact l'échantillon dans lequel la présence de l'acide nucléique cible est suspectée avec une paire d'amorces nucléotidiques dont la position d'hybridation est localisée respectivement
20 du côté 5' et du côté 3' de la région de l'acide nucléique cible dont l'amplification est recherchée, en présence des réactifs nécessaires à la réaction d'amplification ; et

b) détection des acides nucléiques amplifiés.

25 Pour mettre en oeuvre le procédé d'amplification tel que défini ci-dessus, on aura avantageusement recours à l'une quelconque des amorces nucléotidiques décrites ci-avant.

L'invention a en outre pour objet un nécessaire ou kit pour l'amplification d'un acide nucléique selon l'invention, et plus
30 particulièrement tout ou partie d'un acide nucléique de séquences SEQ ID N°1 à SEQ ID N°81, ledit nécessaire ou kit comprenant :

a) un couple d'amorces nucléotidiques conformes à l'invention, dont la position d'hybridation est localisée respectivement du côté 5' et du côté 3' de l'acide nucléique cible dont l'amplification est recherchée ;

b) le cas échéant, les réactifs nécessaires à la réaction d'amplification.

Un tel nécessaire ou kit d'amplification comprendra
5 avantageusement au moins une paire d'amorces nucléotidiques telles que décrites ci-dessus.

Vecteurs recombinants

L'invention est également relative à un vecteur recombinant
10 comprenant un acide nucléique selon l'invention.

Avantageusement, un tel vecteur recombinant comprendra un acide nucléique choisi parmi les acides nucléiques suivants :

a) un acide nucléique codant pour une protéine ayant une séquence en acides aminés choisie dans le groupe des séquences SEQ
15 ID N°129 à SEQ ID N°160 ou un fragment peptidique ou un variant de cette dernière ;

b) un acide nucléique comprenant un polynucléotide choisi dans le groupe constitué des séquences SEQ ID N°1 à SEQ ID N°81, ou un fragment ou un variant de ce dernier ;

20 c) un acide nucléique ayant au moins 80% d'identité en nucléotides avec un acide nucléique choisi dans le groupe constitué des séquences SEQ ID N°1 à SEQ ID N°81 ou un fragment ou un variant de ce dernier ;

d) un acide nucléique hybridant, dans des conditions d'hybridation
25 de forte stringence, avec un acide nucléique de séquences SEQ ID N°1 à SEQ ID N°81, ou un fragment ou un variant de ce dernier.

Par " vecteur " au sens de la présente invention on entendra une molécule d'ADN ou d'ARN circulaire ou linéaire qui est indifféremment
30 sous forme de simple brin ou double brin.

Selon un premier mode de réalisation, un vecteur recombinant selon l'invention est utilisé afin d'amplifier l'acide nucléique qui y est inséré après transformation ou transfection de l'hôte cellulaire désiré.

Selon un second mode de réalisation, il s'agit de vecteurs
35 d'expression comprenant, outre un acide nucléique conforme à

l'invention, des séquences régulatrices permettant d'en diriger la transcription et/ou la traduction.

Selon un mode de réalisation avantageux, un vecteur recombinant selon l'invention comprendra notamment les éléments suivants :

5 (1) des éléments de régulation de l'expression de l'acide nucléique à insérer, tels que des promoteurs et des enhanceurs ;

(2) la séquence codante comprise dans l'acide nucléique conforme à l'invention à insérer dans un tel vecteur, ladite séquence codante étant placée en phase avec les signaux de régulation décrits
10 aux (1) ; et

(3) des séquences d'initiation et d'arrêt de la transcription appropriées.

En outre, les vecteurs recombinants selon l'invention pourront
15 inclure une ou plusieurs origines de réplication chez les hôtes cellulaires dans lesquels leur amplification ou leur expression est recherchée, des marqueurs ou des marqueurs de sélection.

A titre d'exemples, les promoteurs bactériens pourront être les promoteurs LacI, LacZ, les promoteurs de l'ARN polymérase du
20 bactériophage T3 ou T7, les promoteurs PR, ou PL du phage lambda.

Les promoteurs pour cellules eucaryotes comprendront le promoteur de la thymidine kinase du virus HSV ou encore le promoteur de la métallothionéine-L de souris.

De manière générale, pour le choix d'un promoteur adapté,
25 l'homme du métier pourra avantageusement se référer à l'ouvrage de SAMBROOK et al. (1989) précité ou encore aux techniques décrites par FULLER et al. (1996).

Les vecteurs bactériens préférés selon l'invention sont par exemple les vecteurs pBR322(ATCC37017) ou encore des vecteurs tels
30 que pAA223-3 (Pharmacia, Uppsala, Suède), et pGEM1 (Promega Biotech, Madison, WI, ETATS-UNIS).

On peut encore citer d'autres vecteurs commercialisés tels que les vecteurs pQE70, pQE60, pQE9 (Qiagen), psiX174, pBluescript SA, pNH8A, pNH16A, pNH18A, pNH46A, pWLNEO, pSV2CAT, pOG44,
35 pXTI, pSG(Stratagene).

Il peut s'agir également de vecteurs de type *baculovirus* tel que le vecteur pVL1392/1393 (Pharmingén) utilisé pour transfecter les cellules de la lignée Sf9 (ATCC N°CRL 1711) dérivées de *Spodoptera frugiperda*.

5 Il peut encore s'agir de vecteurs adénoviraux tels que l'adénovirus humain de type 2 ou 5.

Un vecteur recombinant selon l'invention peut aussi être un vecteur rétroviral ou encore un vecteur adéno-associé (AAV). De tels vecteurs adéno-associés sont par exemple décrits par FLOTTE et al. (1992), SAMULSKI et al. (1989), ou encore McLAUGHLIN BA et al. (1996).

Cellules hôtes recombinantes

15 L'invention concerne aussi une cellule hôte recombinante comprenant un acide nucléique conforme à l'invention, et plus particulièrement un acide nucléique de séquences SEQ ID N°1 à SEQ ID N°81 ou encore un acide nucléique comprenant tout ou partie de la région codante de ces derniers.

20 Selon un autre aspect, l'invention est également relative à une cellule hôte recombinante comprenant un vecteur recombinant tel que ci-dessus décrit.

Les cellules hôtes préférées selon l'invention sont par exemple les suivantes :

25

a) cellules hôtes procaryotes: souches d'*Escherichia coli* (souche DH5- α), de *Bacillus subtilis*, de *Salmonella typhimurium*, ou encore des souches d'espèces telles que *Pseudomonas*, *Streptomyces* et *Staphylococcus* ;

30

b) cellules hôtes eucaryotes: cellules HeLa (ATCC N°CCL2), cellules Cv 1 (ATCC N°CCL70), cellules COS (ATCC N°CRL 1650), cellules Sf-9 (ATCC N°CRL 1711), cellules CHO (ATCC N°CCL-61) ou encore cellules 3T3 (ATCC N°CRL-6361).

35

Selon un autre aspect, l'invention concerne un polypeptide comprenant une séquence en acides aminés choisie dans le groupe constitué des peptides de séquences SEQ ID N°129 à SEQ ID N°160, ou un fragment peptidique ou un variant de ce dernier.

5 L'invention concerne aussi un polypeptide comprenant au moins 15 acides aminés consécutifs d'une séquence en acides aminés choisie dans le groupe constitué des peptides de séquences SEQ ID N°129 à SEQ ID N°160, ou un fragment peptidique ou un variant de ce dernier

L'invention est également relative à un polypeptide comprenant
10 une séquence en acides aminés ayant au moins 80% d'identité en acides aminés avec une séquence en acides aminés choisie dans le groupe constitué des peptides de séquences SEQ ID N°129 à SEQ ID N°160, ou un fragment peptidique ou un variant de ce dernier.

Avantageusement, fait partie de l'invention un polypeptide ayant
15 au moins 85%, 90%, 95% ou 99% d'identité en acides aminés avec une séquence en acides aminés choisie dans le groupe constitué des peptides de séquences SEQ ID N°129 à SEQ ID N°160, ou un fragment peptidique ou un variant de ce dernier.

20 De préférence, des polypeptides selon l'invention auront une longueur de 15, 18 ou 20 à 25, 35, 40, 50, 70, 80, 100 ou 200 acides aminés consécutifs d'un acide nucléique selon l'invention, en particulier un polypeptide de séquence en acides aminés choisie parmi les séquences SEQ ID N°129 à SEQ ID N°160 .

25 Alternativement, un polypeptide selon l'invention consistera et/ou comprendra les fragments d'une longueur de 15, 18, 20, 25, 35, 40, 50, 100 ou 200 acides aminés consécutifs d'un polypeptide selon l'invention, plus particulièrement d'un polypeptide choisi parmi les séquences SEQ
30 ID N°129 à SEQ ID N°160.

De manière générale, les polypeptides selon la présente invention se présentent sous une forme isolée ou purifiée.

L'invention concerne aussi un polypeptide comprenant des modifications d'acides aminés de 1, 2, 3, 4, 5, 10 à 20 substitutions, additions ou délétions d'un acide aminé par rapport à la séquence en acides aminés d'un polypeptide de séquences SEQ ID N°129 à SEQ ID N°160, ou encore d'un fragment ou d'un variant de ce dernier.

L'invention est également relative à un procédé pour la production de l'un des polypeptides de séquences SEQ ID N°129 à SEQ ID N°160 ou d'un fragment peptidique ou d'un variant de ce dernier, ladite méthode comprenant les étapes de :

- 10 a) insérer un acide nucléique codant pour ledit polypeptide dans un vecteur approprié ;
- b) cultiver, dans un milieu de culture approprié, une cellule hôte préalablement transformée ou transfecter avec le vecteur recombinant de l'étape a) ;
- 15 c) récupérer le milieu de culture conditionné ou lyser la cellule hôte, par exemple par sonication ou par choc osmotique ;
- d) séparer et purifier à partir dudit milieu de culture ou encore à partir des lysats cellulaires obtenus à l'étape c), ledit polypeptide ;
- e) le cas échéant, caractériser le polypeptide recombinant produit.

20

Les peptides selon l'invention peuvent être caractérisés par fixation sur une colonne de chromatographie d'immunoaffinité sur laquelle les anticorps dirigés contre ce polypeptide ou contre un fragment ou un variant de ce dernier ont été préalablement immobilisés.

25 Selon un autre aspect, un polypeptide recombinant selon l'invention peut être purifié par passage sur une série appropriée de colonnes de chromatographie, selon les méthodes connues de l'homme de l'art et décrites par exemple dans F.Ausubel et al (1999).

Un polypeptide selon l'invention peut être également préparé par les techniques classiques de synthèse chimique indifféremment en solution homogène ou phase solide.

A titre illustratif, un polypeptide selon l'invention pourra être préparé par la technique ou en solution homogène décrite par HOUBENWEYL (1974) ou encore la technique de synthèse en phase solide décrite par MERRIFIELD (1965a; 1965b).

35

Font également partie de l'invention des polypeptides dits "homologues" à l'un quelconque des polypeptides de séquences d'acides aminés SEQ ID N°129 à SEQ ID N°160, ou de leurs fragments ou variants.

5 De tels polypeptides homologues ont des séquences d'acides aminés possédant une ou plusieurs substitutions d'un acide aminé par un acide aminé équivalent, par rapport aux polypeptides de référence.

On entendra par acide aminé équivalent selon la présente invention, par exemple remplacement d'un résidu sous la forme L par un
10 résidu sous la forme D ou encore le remplacement d'un acide glutamique (E) par un acide pyro-glutamique selon des techniques bien connues de l'homme du métier. A titre illustratif, la synthèse de peptide contenant au moins un résidu sous la forme D est décrite par KOCH (1977).

15 Selon un autre aspect, sont également considérés comme des acides aminés équivalents deux acides aminés appartenant à la même classe, c'est-à-dire deux acides aminés acide, basique, non polaire ou encore polaire non chargé.

Font également partis de l'invention des polypeptides comprenant
20 au moins une liaison non peptidique telle qu'une liaison rétro-inverso (NHCO), une liaison carba (CH₂CH₂) ou encore une liaison cétométhylène (CO-CH₂).

De préférence, les polypeptides selon l'invention comprenant une ou plusieurs additions, délétions, substitutions d'au moins un acide
25 aminé conserveront leur capacité à être reconnus par des anticorps dirigés contre les polypeptides non modifiés.

Anticorps

30

Les polypeptides selon l'invention, en particulier les polypeptides de séquences en acides aminés SEQ ID N°129 à SEQ ID N°160 ou les fragments et les variants de ces derniers ainsi que les peptides homologues peuvent être utilisés pour la préparation d'anticorps.

35

Par " anticorps " au sens de la présente invention, on entendra notamment des anticorps polyclonaux ou monoclonaux ou des fragments (par exemple les fragments F (ab)'₂, Fab) ou encore tout polypeptide comprenant un domaine de l'anticorps initial reconnaissant le polypeptide ou le fragment de polypeptide cible selon l'invention .

Des anticorps monoclonaux peuvent être préparés à partir d'hybridomes selon la technique décrite par KOHLER et MILSTEIN (1975).

La présente invention concerne également des anticorps dirigés contre un polypeptide tel que décrit ci-dessus ou un fragment ou un variant de ce dernier, tels que produits dans la technique du trioma ou encore la technique d'hybridome décrite par KOZBOR et al. (1983).

L'invention a également trait à des fragments d'anticorps simple chaîne Fv (ScFv) tels que décrits dans le brevet US N° 4,946,778 ou encore par MARTINEAU et al. (1998).

Les anticorps selon l'invention comprennent également des fragments d'anticorps obtenus à l'aide de banques de phages RIDDER et al., (1995) ou encore des anticorps humanisés REIMANN et al. (1997); LEGER et al., (1997).

Les préparations d'anticorps selon l'invention sont utiles dans des tests de détection immunologiques destinés à l'identification de la présence et/ou de la quantité d'antigènes présents dans un échantillon.

Un anticorps selon l'invention pourra comprendre en outre un marqueur détectable isotopique ou non-isotopique, par exemple fluorescent ou encore être couplé à une molécule telle que la biotine, selon des techniques bien connues de l'homme du métier.

Ainsi, la mention a en outre pour objet un procédé pour détecter la présence d'un polypeptide conforme à l'invention dans un échantillon, ledit procédé comprenant les étapes de :

a) mettre en contact l'échantillon à tester avec un anticorps tel que décrit ci-dessus ;

b) détecter le complexe antigène/anticorps formé.

L'invention est également relative à un nécessaire ou kit de diagnostic ou pour la détection de la présence d'un polypeptide conforme à l'invention dans un échantillon, ledit nécessaire comprenant :

- a) un anticorps tel que défini ci-dessus ;
- 5 b) un réactif permettant la détection des complexes antigène/anticorps formés.

Méthode de criblage d'une molécule ou d'une substance se fixant sur un polypeptide selon l'invention.

- 10 Un polypeptide selon l'invention peut être utilisé pour cribler des molécules se fixant sur celui-ci.

La fixation du polypeptide avec la molécule ou substance peut activer (molécule agoniste) ou inhiber (molécule antagoniste) l'activité dudit polypeptide.

- 15 De telles molécules capables de se fixer sur l'un quelconque des polypeptides selon l'invention comprennent des anticorps, des oligonucléotides, d'autres protéines et de manière générale des petites molécules de toutes natures.

- 20 Dans un tel test de criblage, on peut simplement mettre en évidence la fixation de la molécule candidate aux polypeptides, l'un des deux partenaires étant marqué par un composé détectable (polypeptide d'intérêt ou molécule candidate), la visualisation du complexe polypeptide/molécule candidate étant alors visualisée par détection du marqueur détectable, après élimination des molécules candidates non
- 25 liées spécifiquement.

- A titre d'exemple, un test de criblage d'une molécule candidate capable de se fixer sur un polypeptide selon l'invention pourra comprendre avantageusement une première étape au cours de laquelle le polypeptide d'intérêt ou la molécule candidate est immobilisé sur un
- 30 support, une seconde étape au cours de laquelle le second partenaire (molécule candidate ou polypeptide d'intérêt) est mis en présence du premier composé préalablement immobilisé sur le support, une troisième étape au cours de laquelle un ou plusieurs lavages sont effectués dans des conditions appropriées à l'élimination des composés n'étant pas liés
- 35 spécifiquement, et enfin une quatrième étape au cours de laquelle le

complexe éventuellement formé entre le polypeptide d'intérêt et la molécule candidate est détecté.

Dans le mode de réalisation du test de criblage selon lequel la molécule candidate est préalablement immobilisée sur un support, puis
5 mise en présence du polypeptide d'intérêt selon l'invention, la détection du complexe formé par la molécule candidate et le polypeptide d'intérêt selon l'invention pourra être avantageusement réalisée à l'aide d'un anticorps tel que décrit ci-dessus.

Dans un autre mode de réalisation du test de criblage selon lequel
10 c'est le polypeptide d'intérêt selon l'invention qui est préalablement immobilisé sur un support, la molécule candidate sera avantageusement marquée à l'aide d'un marqueur détectable préalablement à sa mise en contact avec le polypeptide d'intérêt immobilisé.

Un tel marqueur détectable peut être radioactif ou non radioactif,
15 par exemple fluorescent ou correspondre à un ligand pour un troisième partenaire utilisé pour la détection comme une molécule de biotine.

En conséquence, l'invention a également pour objet un procédé de criblage d'une molécule ou d'une substance candidate interagissant
20 avec un polypeptide selon l'invention, ladite méthode comprenant les étapes de :

- a) mettre en contact un polypeptide conforme à l'invention avec la substance ou molécule candidate à tester ;
- b) détecter les complexes éventuellement formés entre ledit
25 polypeptide et ladite substance ou molécule candidate.

L'invention concerne également un nécessaire ou kit pour le criblage d'une molécule ou d'une substance candidate interagissant avec un polypeptide selon l'invention, ledit nécessaire comprenant :

- 30 a) un polypeptide conforme à l'invention ;
- b) le cas échéant, des moyens nécessaires à la détection du complexe formé entre ledit polypeptide et la molécule ou substance candidate.

La présente invention est en outre illustrée, sans pour autant être limitée, par les exemples suivants :

EXEMPLES

Exemple 1: Distribution tissulaire des transcrits selon l'invention

Le profil d'expression des polynucleotides selon la présente invention est déterminé selon les protocoles d'analyse de Northern blot et de transcription inverse couplée à la PCR décrits notamment par
10 Sambrook et al (ref. CSH Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). " Molecular Cloning : A Laboratory Manual, " 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.).

15 Par exemple, dans le cas d'une analyse par transcription inverse, un couple d'amorces synthétisées à partir de l'une quelconque des séquences nucléotidiques des transcrits SEQ ID N°1 à SEQ ID N°81 est utilisé pour détecter l'ADNc correspondant.

20 La réaction de polymérase en chaîne (PCR) est réalisée sur des matrices d'ADNc correspondant à des ARNm polyA⁺ (Clontech) rétrotranscrits. La transcription inverse en ADNc est réalisée avec l'enzyme SUPERScript II (GibcoBRL, Life Technologies) selon les conditions décrites par le fabricant.

25 La réaction de polymérase en chaîne est réalisée selon des conditions standard, dans 20 µl de mélange réactionnel avec 25 ng de la préparation d'ADNc. Le mélange réactionnel est composé de 400 µM de chacun des dNTP, de 2 unités de *Thermus aquaticus* (Taq) ADN polymérase (Ampli Taq Gold ; Perkin Elmer), de 0,5 µM de chaque
30 amorce, de 2,5 mM MgCl₂, et de tampon PCR. Trente quatre cycles de PCR (dénaturation 30 s à 94 °C, hybridation de 30 s décomposé

comme suit lors des 34 cycles : 64°C 2 cycles, 61°C 2 cycles, 58°C 2 cycles et 55°C 28 cycles et une élongation d'une minute par kilobase à 72°C) sont réalisés après une première étape de dénaturation à 94°C durant 10 min dans un thermocycler Perkin Elmer 9700. Les réactions de PCR sont visualisées sur gel d'agarose par électrophorèse. Les fragments d'ADNc obtenus peuvent être utilisés comme sondes pour une analyse par Northern blot et peuvent également être utilisés pour la détermination exacte de la séquence polynucléotidique.

10 Dans le cas d'une analyse par Northern Blot, une sonde d'ADNc produite comme décrit ci-dessus est marquée au ^{32}P grâce au système de marquage d'ADN High Prime (Boehringer) selon les instructions indiquées par le fabricant. Après marquage, la sonde est purifiée sur une microcolonne de Sephadex G50 (Pharmacia) selon les instructions indiquées par le fabricant. La sonde marquée et purifiée est alors utilisée pour la détection de l'expression des ARNm dans différents tissus.

Le Northern blot contenant des échantillons d'ARN de différents tissus humains ((Multiple Tissue Northern , MTN, Clontech) Blot 2, référence 77759-1) est hybridé avec la sonde marquée.

20 Le protocole suivi pour les hybridations et lavages peut être soit directement celui décrit par le fabricant (Manuel d'utilisation PT1200-1) soit une adaptation de ce protocole en utilisant les méthodes connues de l'homme de l'art et décrites par exemple dans F.AUSUBEL et al (1999).

On pourra ainsi faire varier par exemple les températures de préhybridation et d'hybridation en présence de formamide.

Par exemple on pourra utiliser le protocole suivant :

1- Compétition des membranes et pré-hybridation:

30

- Mélanger : 40µl ADN sperme de saumon (10mg/ml)
+ 40 µl ADN placentaire humain (10mg/ml)

- Dénaturer 5 mn à 96°C, puis plonger dans la glace le mélange.

5 - Oter le SSC 2X et verser 4 ml de mix formamide dans le tube d'hybridation contenant les membranes.

- Ajouter le mélange des deux ADN dénaturés.

10 - Incubation à 42°C pendant 5 à 6 heures, avec rotation.

2- Compétition de la sonde marquée :

15 - Ajouter à la sonde marquée et purifiée 10 à 50 µl ADN Cot I, selon la quantité de repeats.

- Dénaturer 7 à 10 mn à 95°C.

- Incuber à 65°C pendant 2 à 5 heures.

20 **3- HYBRIDATION :**

- Oter le mix de pré hybridation.

25 - Mélanger 40 µl ADN sperme de saumon + 40 µl ADN placentaire humain ; dénaturer 5 mn à 96°C, puis plonger dans la glace.

- Ajouter dans le tube d'hybridation 4 ml de mix formamide, le mélange des deux ADN et la sonde marquée/ADN.Cot I dénaturée.

30 - Incuber 15 à 20 heures à 42°C, avec rotation.

4- Lavages :

35 - Un lavage à température ambiante dans du SSC 2X, pour rincer.

- 2 fois 5 minutes à température ambiante SSC 2X et SDS 0,1% à 65°C.

- 2 fois 15 minutes à 65°C SSC 1X et SDS 0,1% à 65°C.

40

Après hybridation et lavage, le blot est analysé après une nuit d'exposition au contact d'un écran au phosphore révélé à l'aide du Storm (Molecular Dynamics, Sunnyvale, CA).

**Exemple 2 : Obtention de fragments d'ADNc complets
correspondant aux transcrits selon l'invention**

- Différentes approches peuvent être utilisées pour isoler l'ADNc
5 correspondant à un des clones particulier parmi les séquences SEQ ID
N° 1 à SEQ ID N° 81.

Par exemple un clone complet peut être directement isolé par
hybridation en criblant une banque d'ADNc au moyen d'une sonde
polynucléotidique spécifique de la séquence du gène d'intérêt.

10 En particulier une sonde spécifique de 30-40 nucléotides est
synthétisée en utilisant un synthétiseur de marque Applied
Biosystem/Perkin Elmer selon la séquence choisie.

L'oligonucléotide obtenu est radiomarké, par exemple au ^{32}P - γ -
ATP en utilisant la T4 polynucléotide kinase et est purifié selon les
15 méthodes usuelles (e.g. Maniatis et al. Molecular cloning : A Laboratory
Manual, Cold Spring Harbor Press, Cold Spring, NY 1982 ou encore
F.Ausubel et al. (Current Protocols in Molecular Biology, J.Wiley and
Sons Eds, 1999).

20 La banque de clones contenant l'ADNc que l'on veut cribler est
étalée sur milieu de culture en boîte de Pétri (1.5% agar) contenant les
antibiotiques appropriés selon les méthodes usuelles citées ci dessus
(F.Ausubel et al.). Les colonies ainsi produites après incubation sont
transférées sur filtres de nitrocellulose et criblées au moyen de la sonde
25 nucléotidique radiomarkée, selon les méthodes usuelles et les colonies
hybridant avec la sonde sont isolées et sous clonées.

L'ADN des clones ainsi repéré est préparé et analysé par
séquençage. Les clones contenant les fragments correspondant à
30 l'ADNc complet sont purifiés et reclés dans le vecteur pcDNA3 selon

les protocoles connus de l'homme de l'art et présentés par exemple dans F.Ausubel et al (1999) .

Différentes méthodes sont connues pour identifier les extrémités 5' et 3' du cDNA correspondant aux gènes décrits dans la présente demande. Ces méthodes incluent mais ne se limitent pas au clonage par hybridation, au clonage utilisant des protocoles similaires ou identiques à la 3' ou 5' RACE-PCR (Rapid Amplification of cDNA End-PCR) qui sont bien connues de l'homme de l'art.

10

Par exemple, on pourra utiliser le kit commercialisé par la société Clontech (Marathon Ready™ cDNA kit , protocole référencé PT1156-1) ou alternativement une méthode similaire à la 5'RACE est disponible pour caractériser l'extrémité 5' manquante d'un cDNA (Fromont-Racine et al. Nucleic Acid Res.21(7):1683-1684 (1993)). En bref, un oligonucléotide d'ARN est ligaturé à l'extrémité 5' d'une population d'ARNm. Après retrotranscription en cDNA, un jeu d'amorces spécifiques respectivement de l'adaptateur ligaturé en 5' et d'une séquence située en 3' du gène d'intérêt est utilisé en PCR pour amplifier la portion 5' du cDNA recherché. Le fragment amplifié est ensuite utilisé pour reconstruire l'ADNc complet.

Exemple 3 : Analyse du profil d'expression génique pour la maladie de Tangier

25

La vérification de la perte d'expression du gène candidat entraînant le phénotype cellulaire de Tangier peut-être déterminé par hybridation de ces séquences avec des sondes correspondant aux

ARNm provenant de fibroblastes de sujets atteints ou non de la maladie, selon les méthodes décrites ci-dessous :

1. Préparation des ARN totaux, des ARNm poly(A)⁺ et de sondes de cDNA

Les ARN totaux sont obtenus à partir de cultures cellulaires des fibroblastes de sujets normaux ou bien atteints de la maladie de Tangier par la méthode à l'isothiocyanate de guanidine (Chomczynski & Sacchi, 1987). Les ARNm poly(A)⁺ sont obtenus par chromatographie d'affinité sur colonnes d'oligo(dT)-cellulose (Sambrook et al., 1989) et les cDNA utilisés comme sondes sont obtenus par RT-PCR (DeRisi et al., 1997) avec des oligonucléotides marqués avec un produit fluorescent (Amersham Pharmacia Biotech ; CyDyeTM).

2. Hybridation et détection des niveaux d'expressions

Les membranes de verre contenant les séquences présentées dans cette demande de brevet, correspondant au gène Tangier sont hybridées avec les sondes de cDNA, obtenues à partir des fibroblastes (Iyer et al., 1999). L'utilisation du système Amersham/molecular Dynamics (Avalanche MicroscannerTM) permet la quantification des expressions des produits de séquences sur le type cellulaire sain ou affecté.

Exemple 4 : Construction du vecteur d'expression dans des cellules de mammifères

Le gène d'intérêt peut être exprimé dans des cellules de mammifères. Un vecteur d'expression eukaryote typique contient un promoteur qui permet l'initiation de la transcription de l'ARNm, une

séquence codant pour la protéine, et les signaux requis pour la terminaison de la transcription et pour la polyadénylation du transcrit. Il contient aussi des signaux supplémentaires comme des enhancers, la séquence (de) Kozak et des séquences nécessaires pour l'épissage de l'ARNm. Une transcription efficace est obtenue avec les éléments early et late des promoteurs du virus SV40, les LTR rétroviraux ou le promoteur early du virus CMV. Cependant des éléments cellulaires comme le promoteur de l'actine peuvent aussi être employés. De nombreux vecteurs d'expression peuvent être employés pour mettre en œuvre la présente invention comme le vecteur pcDNA3.

Exemple 5 : Production des polypeptides

Le polypeptide correspondant au transcrit partiel du gène GS N°XX ou au cDNA complet décrit dans l'Exemple 2 (clonage du cDNA complet) peut être facilement produit dans un système d'expression bactérienne, de cellules d'insectes en utilisant les vecteurs baculovirus ou encore dans des cellules de mammifères avec ou sans les vecteurs du virus de la vaccine. Toutes les méthodes sont aujourd'hui largement décrites et connues de l'homme de l'art. On en trouvera par exemple une description détaillée dans F.Ausubel et al. (1999).

Exemple 6 : Production d'un anticorps dérivé d'un polypeptide

Les anticorps dans la présente invention peuvent être préparés par différentes méthodes (Current Protocols In Molecular Biology Volume 1 edited by Frederick M. Ausubel, Roger Brent, Robert E. Kingston, David D. Moore, J.G. Seidman, John A. Smith, Kevin Struhl - Massachusetts General Hospital Harvard Medical School, chapitre 11).

Par exemple, les cellules exprimant un polypeptide de la présente invention sont injectées dans un animal afin d'induire la production de serum contenant les anticorps. Dans une des méthodes décrites, les protéines sont préparées et purifiées afin d'éviter des contaminations.

5 Une telle préparation est alors introduite dans l'animal dans le but de produire des antisera polyclonaux de plus grande activité.

Dans la méthode préférée, les anticorps de la présente invention sont des anticorps monoclonaux. De tels anticorps monoclonaux peuvent être préparés en utilisant la technique d'hybridome. (Köhler et al, Nature 256 :495 (1975) ; Köhler et al, Eur. J. Immunol. 6 :511 (1976) ;

10 Köhler et al, Eur. J. Immunol. 6:292 (1976) ; Hammeling et al., in : Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., pp. 563-681 51981). En général, de telles méthodes impliquent d'immuniser l'animal (préférentiellement une souris) avec un polypeptide ou, mieux

15 encore, avec une cellule exprimant le polypeptide. Ces cellules peuvent être mises en culture dans un milieu de culture tissulaire adéquat. Cependant, il est préférable de cultiver les cellules dans un milieu Eagle (Earle modifié) supplémenté avec 10% de serum bovin foetal (inactivé à 56°C) et additionné d'environ 10 g /l d'acides aminés non essentiels, de

20 1000 U/ml de pénicilline et d'environ 100 µg/ml de streptomycine.

Les splénocytes de ces souris sont extraits et fusionnés avec une lignée cellulaire de myelome adéquate. Cependant, il est préférable d'utiliser la lignée cellulaire de myelome parentale (SP2O) disponible à l'ATCC. Après fusion, les cellules d'hybridomes résultantes sont

25 sélectivement maintenues en milieu HAT puis clonées par dilution limite comme décrit par Wands et al. (Gastroentérolgy 80:225-232 (1981)). Les cellules d'hybridomes obtenues après une telle sélection sont testées afin d'identifier les clones sécrétant des anticorps capables de se fixer au polypeptide.

30 D'autre part, d'autres anticorps capables de se fixer au polypeptide peuvent être produits selon une procédure en 2 étapes

utilisant des anticorps anti-idiotypique une telle méthode est fondée sur le fait que les anticorps sont eux-mêmes des antigènes et en conséquence il est possible d'obtenir un anticorps reconnaissant un autre anticorps. Selon cette méthode, les anticorps spécifiques de la protéine sont utilisés pour immuniser un animal, préférentiellement une souris. Les splénocytes de cet animal sont ensuite utilisés pour produire des cellules hybridomes, et ces dernières sont criblées pour identifier les clones qui produisent un anticorps dont la capacité à se fixer au complexe protéine-anticorps spécifique peut-être bloqué par le polypeptide. Ces anticorps peuvent être utilisés pour immuniser un animal afin d'induire la formation en plus grande quantité d'anticorps spécifiques de la protéine.

Il serait apprécié que Fab et F(ab')₂ et les autres fragments des anticorps de la présente invention puissent être utilisés selon les méthodes décrites ici. De tels fragments sont typiquement produits par clivage protéolytique à l'aide d'enzymes telles que la Papaine (pour produire les fragments Fab) ou la Pepsine (pour produire les fragments F(ab')₂). Sinon, les fragments sécrétés reconnaissant la protéine peuvent être produits en appliquant la technologie de l'ADN recombinant ou de la chimie de synthèse.

Pour l'utilisation in vivo d'anticorps chez l'homme il serait préférable d'utiliser des anticorps monoclonaux chimériques "humanisés". De tels anticorps peuvent être produits en utilisant des constructions génétiques dérivés de cellules d'hybridomes produisant les anticorps monoclonaux décrits ci-dessus. Les méthodes pour produire les anticorps chimériques sont connus par l'homme de l'art. (Pour revue, voir : Morrison, Science 229 :1202 (1985) ; Oi et al., Biotechnique 4 :214 (1986) ; Cabilly et al., US patent n°4,816,567 ; Taniguchi et al., EP 171496 ; Morrison et al., EP 173494 ; Neuberger et al., WO 8601533 ; Robinson et al., WO 8702671 ; Boulianne et al ; Nature 312 :643 (1984) ; Neuberger et al., Nature 314 : 268 (1985)).

Exemple 7 : Correction du phénotype cellulaire de la maladie de Tangier

5

La maladie de Tangier est caractérisée par un catabolisme accéléré des particules lipoprotéiques de haute densité (HDL) et une accumulation de cholestérol dans les tissus. Notamment, les fibroblastes de peau des patients atteints de la maladie de Tangier ont une capacité
10 réduite à éliminer leur contenu en cholestérol par le processus d'efflux de cholestérol assuré par l'apolipoprotéine A-I (apoA-I), protéine majeure des HDL (Francis et al., 1995). Cette caractéristique correspondant à une perte de fonction est aussi retrouvée dans d'autres cellules fibroblastiques de patients atteints de déficit familial en HDL (Marcil et
15 al., 1999).

La correction du phénotype des fibroblastes de Tangier peut être assurée par la transfection des ADNc complets correspondant aux séquences proposées, dans lesdites cellules. L'ADNc est inséré dans un
20 vecteur d'expression qui est ensuite transfecté selon les méthodes décrites ci-dessous :

1. Préparation des cultures fibroblastiques de sujets normaux et de sujets atteints de la maladie de Tangier

25 Les fibroblastes primaires de peau humaine sont obtenus par la mise en culture de biopsie de peau provenant de l'avant bras. Ces biopsies sont effectuées sur des patients atteints de la maladie de Tangier ayant les particularités cliniques et biochimiques des "homozygotes", c'est à dire des amygdales oranges, des
30 concentrations plasmatiques d'apoA-I et de cholestérol-HDL inférieur au

5^{ème} percentile. Les lignées de fibroblastes normaux sont obtenus chez l'American Type Culture Collection (Rockville, MD). Les fibroblastes sont cultivés dans un milieu EMMEM (Eagle-modified minimum essential medium ; GIBCO) complété par 10% de sérum de veau foetal, de la glutamine à 2 mM, 100 UI/ml de pénicilline et 100 µg/ml de streptomycine (milieu désigné par EMMEM10). En vue de la réalisation de l'étude de l'efflux de cholestérol, ces cellules sont pré-chargées en cholestérol par incubation de 24 heures avec 50µg/ml de cholestérol dans le milieu décrit ci-dessus sans sérum de veau mais contenant 2 mg/ml d'albumine bovine (BSA, fraction V).

2. Etude de l'efflux de cholestérol

Les fibroblastes pré-chargés en cholestérol à confluence sur des plaques à 24 puits sont incubés dans le milieu EMMEM10 et 1 µCi/ml de 1,2-³H-cholestérol (50 Ci/mmol ; Dupont ; Wilmington, DE) durant 48 heures. Environ 100 000 coups par minute sont obtenus par puits ou 1000 coups par minute par µg de protéine cellulaire. Les cellules sont lavées trois fois avec du milieu EMMEM/BSA, et incubées avec ce milieu durant 24 heures avant de transfecter le gène d'intérêt et de démarrer l'efflux par ajout de 10 µg/ml de protéoliposome contenant l'apoA-I en milieu EMMEM/BSA. Ces protéoliposomes sont préparés par sonication de phosphatidylcholine et d'apoA-I humaine purifiée (Jonas, 1986). La transfection cellulaire s'effectue par la technique de précipitation de phosphate de calcium (Sambrook et al., 1989). Après la période d'efflux, en général 20 heures, le milieu est collecté, centrifugé (1000 g, 5 min), et la radioactivité déterminée par comptage en scintillation liquide. La radioactivité résiduelle dans les cellules est aussi déterminée sur la nuit après extraction des lipides dans l'isopropanol. Le pourcentage d'efflux est calculé en divisant la radioactivité mesurée dans le surnageant par la somme des radioactivités mesurées, dans le surnageant et l'extrait cellulaire. Un contrôle interne est réalisé par transfection d'un gène

marqueur et l'incubation sur 24 heures avec un milieu EMMEM/BSA sans protéoliposome contenant l'apoA-I. L'efflux de cholestérol cellulaire à partir de fibroblastes normaux et transfectés par un gène témoin correspondent à $6 \pm 2\%$ alors que celui obtenu à partir de fibroblastes atteints de la maladie de Tangier et transfectés par ce gène témoin est inférieur à 1%. En revanche la transfection des fibroblastes atteints de la maladie de Tangier par un plasmide correspondant aux gènes proposés dans ce dossier permettrait de restaurer la capacité de ces cellules à éliminer leur excès de cholestérol à un niveau correspondant à celui de fibroblastes normaux.

Exemple 8 : Isolement des clones génomiques correspondant à un polynucléotide

L'isolement des clones génomiques correspondant à un transcrit est réalisé par criblage de la banque de BAC d'ADN génomique humain (par exemple celle fournie par Mel Simon, CalTech., ref : Kim et al. Genomics (1996), 34 :213-218)) par PCR avec des amorces spécifiques de la séquence d'ADNc correspondant aux séquences des transcrits SEQ ID N°1 à SEQ ID N°81 selon l'invention, selon la méthode décrite dans l'Exemple 1.

Exemple 9 : Détermination de polymorphisme/mutation de l'un des gènes correspondant aux transcrits selon l'invention

La détection de polymorphismes et ou de mutations dans les séquences des transcrits peut être réalisée selon différents protocoles. La méthode de choix est le séquençage direct.

Dans le cas d'un transcrit où la structure du gène correspondant est inconnue ou partiellement connue il est nécessaire de déterminer précisément sa structure intron-exon ainsi que la séquence génomique du gène correspondant. Il s'agit donc dans un premier temps d'isoler le

ou les clones de BAC d'ADN génomique correspondant au transcrit étudié selon la méthode décrite dans l'exemple 8, de séquencer l'insert du ou des clones correspondants et de déterminer la structure intron-exon en comparant la séquence de l'ADNc à celle de l'ADN génomique obtenu.

La technique de détection de mutation par séquençage direct consiste à comparer les séquences génomiques du gène correspondant à l'ADNc SEQ ID N° 1 à SEQ ID N°81 obtenues à partir d'au moins 8 individus (4 individus affectés par la pathologie étudiée et 4 individus non affectés). Les divergences de séquence constituent des polymorphismes. Tous ceux modifiant la séquence en acides aminés de la protéine sauvage sont des mutations susceptibles d'affecter la fonction de ladite protéine qu'il est intéressant de considérer plus particulièrement dans les études d'association cas/témoin décrite dans l'Exemple 8.

Exemple 10 : Identification du gène causal par la mutation causale ou une différence transcriptionnelle

Parmi les mutations identifiées selon la méthode décrite dans l'Exemple 9, toutes celles associées au phénotype malade sont susceptibles d'être causales. La validation de ces résultats est faite en séquençant le gène chez tous les individus atteints et leurs apparentés (dont l'ADN est disponible).

D'autre part, la réalisation de Northern blot ou RT-PCR, selon la méthode décrite dans l'Exemple 1, à partir d'ARN spécifique d'individus atteints et non-atteints permet de détecter des variations notables du niveau d'expression du gène étudié, en particulier une absence de transcription du gène.

TABLEAU I
BREVE DESCRIPTION DES SEQUENCES SELON L'INVENTION

N° de Séquence	Désignation
1	ADNc correspondant au gène GS9002S31
2	1 ^{er} ADNc correspondant au gène GS910331
3	2 ^e ADNc correspondant au gène GS910331
4	ADNc correspondant au gène GS914554
5	1 ^{er} ADNc correspondant au gène GS914739
6	2 ^e ADNc correspondant au gène GS914739
7	ADNc correspondant au gène GS915574
8	ADNc correspondant au gène GS930321
9	ADNc correspondant au gène GS931311
10	ADNc correspondant au gèneGS934660
11	ADNc correspondant au gène GS938315
12	1 ^{er} ADNc correspondant au gène GS93953
13	2 ^e ADNc correspondant au gène GS93953
14	1 ^{er} ADNc correspondant au gèneGS939874
15	2 ^e ADNc correspondant au gèneGS939874
16	ADNc correspondant au gène GS911370
17	1 ^{er} ADNc correspondant au gèneGS913920
18	2 ^e ADNc correspondant au gèneGS913920
19	1 ^{er} ADNc correspondant au gèneGS91437
20	2 ^e ADNc correspondant au gèneGS91437
21	1 ^{er} ADNc correspondant au gène GS91507
22	2 ^e ADNc correspondant au gène GS91507
23	1 ^{er} ADNc correspondant au gèneGS915231
24	2 ^e ADNc correspondant au gèneGS915231
25	1 ^{er} ADNc correspondant au gène GS915528
26	2 ^e ADNc correspondant au gène GS915528
27	1 ^{er} ADNc correspondant au gène GS99817

TABLEAU I (suite 1)
BREVE DESCRIPTION DES SEQUENCES SELON L'INVENTION

N° de Séquence	Désignation
28	2è ADNc correspondant au gène GS99817
29	ADNc correspondant au gène GS916229
30	1 ^{er} ADNc correspondant au gène GS92544
31	2è ADNc correspondant au gène GS92544
32	3è ADNc correspondant au gène GS92544
33	1 ^{er} ADNc correspondant au gène GS930284
34	2è ADNc correspondant au gène GS930284
35	ADNc correspondant au gène GS93382
36	ADNc correspondant au gène GS946300
37	1 ^{er} ADNc correspondant au gène GS937345
38	2è ADNc correspondant au gène GS937345
39	3è ADNc correspondant au gène GS937345
40	1 ^{er} ADNc correspondant au gène GS99556
41	2è ADNc correspondant au gène GS99556
42	1 ^{er} ADNc correspondant au gène GS96663
43	2è ADNc correspondant au gène GS96663
44	3è ADNc correspondant au gène GS96663
45	4è ADNc correspondant au gène GS96663
46	5è ADNc correspondant au gène GS96663
47	6è ADNc correspondant au gène GS96663
48	1 ^{er} ADNc correspondant au gène GS941675
49	2è ADNc correspondant au gène GS941675
50	3è ADNc correspondant au gène GS941675
51	1 ^{er} ADNc correspondant au gène GS929341
52	2è ADNc correspondant au gène GS929341

TABLEAU I (suite 2)
BREVE DESCRIPTION DES SEQUENCES SELON L'INVENTION

N° de Séquence	Désignation
53	3 ^è ADNc correspondant au gène GS929341
54	1 ^{er} ADNc correspondant au gène GS915742
55	2 ^è ADNc correspondant au gène GS915742
56	3 ^è ADNc correspondant au gène GS915742
57	1 ^{er} ADNc correspondant au gène GS913018
58	2 ^è ADNc correspondant au gène GS913018
59	3 ^è ADNc correspondant au gène GS913018
60	1 ^{er} ADNc correspondant au gène GS911742
61	2 ^è ADNc correspondant au gène GS911742
62	3 ^è ADNc correspondant au gène GS911742
63	1 ^{er} ADNc correspondant au gène GS98601
64	2 ^è ADNc correspondant au gène GS98601
65	3 ^è ADNc correspondant au gène GS98601
66	1 ^{er} ADNc correspondant au gène GS94852
67	2 ^è ADNc correspondant au gène GS94852
68	3 ^è ADNc correspondant au gène GS94852
69	1 ^{er} ADNc correspondant au gène GS935135
70	2 ^è ADNc correspondant au gène GS935135
71	3 ^è ADNc correspondant au gène GS935135
72	1 ^{er} ADNc correspondant au gène GS914669
73	2 ^è ADNc correspondant au gène GS914669
74	3 ^è ADNc correspondant au gène GS914669
75	1 ^{er} ADNc correspondant au gène GS913839
76	2 ^è ADNc correspondant au gène GS913839
77	3 ^è ADNc correspondant au gène GS913839
78	1 ^{er} ADNc correspondant au gène GS912639
79	2 ^è ADNc correspondant au gène GS912639

TABLEAU I (suite 3)
BREVE DESCRIPTION DES SEQUENCES SELON L'INVENTION

N° de Séquence	Désignation
80	3è ADNc correspondant au gène GS912639
81	ADNc correspondant au gène GS933630
82	Sonde pour la séquence SEQ ID NO 2
83	Sonde pour la séquence SEQ ID NO 2
84	Sonde pour la séquence SEQ ID NO 4
85	Sonde pour la séquence SEQ ID NO 5
86	Sonde pour la séquence SEQ ID NO 10
87	Sonde pour la séquence SEQ ID NO 12
88	Sonde pour la séquence SEQ ID No 16
89	Sonde pour la séquence SEQ ID NO 16
90	Sonde pour la séquence SEQ ID NO 21
91	Sonde pour la séquence SEQ ID NO 23
92	Sonde pour la séquence SEQ ID NO 25
93	Sonde pour la séquence SEQ ID NO 27
94	Sonde pour la séquence SEQ ID NO 30
95	Sonde pour la séquence SEQ ID NO 33
96	Sonde pour la séquence SEQ ID NO 33
97	Sonde pour la séquence SEQ ID NO 35
98	Sonde pour la séquence SEQ ID NO 40
99	Sonde pour la séquence SEQ ID NO 40
100	Sonde pour la séquence SEQ ID NO 40
101	Sonde pour la séquence SEQ ID NO 60
102	Amorce dans SEQ ID NO 17
103	Amorce dans SEQ ID NO 17
104	Amorce dans SEQ ID NO 37
105	Amorce dans SEQ ID NO 37
106	Amorce dans SEQ ID NO 42

TABLEAU I (suite 4)
BREVE DESCRIPTION DES SEQUENCES SELON L'INVENTION

N° de Séquence	Désignation
107	Amorce dans SEQ ID NO 42
108	Amorce dans SEQ ID NO 43
109	Amorce dans SEQ ID NO 45
110	Amorce dans SEQ ID NO 45
111	Amorce dans SEQ ID NO 57
112	Amorce dans SEQ ID NO 57
113	Amorce dans SEQ ID NO 58
114	Amorce dans SEQ ID NO 58
115	Amorce dans SEQ ID NO 66
116	Amorce dans SEQ ID NO 66
117	Amorce dans SEQ ID NO 67
118	Amorce dans SEQ ID NO 67
119	Amorce dans SEQ ID NO 69
120	Amorce dans SEQ ID NO 70
121	Amorce dans SEQ ID NO 72
122	Amorce dans SEQ ID NO 72
123	Amorce dans SEQ ID NO 73
124	Amorce dans SEQ ID NO 73
125	Amorce dans SEQ ID NO 75
126	Amorce dans SEQ ID NO 76
127	Amorce dans SEQ ID NO 78
128	Amorce dans SEQ ID NO 79
129	Polypeptide codé par la séquence SEQ ID NO 6
130	Polypeptide codé par la séquence SEQ ID NO 9
131	Polypeptide codé par la séquence SEQ ID NO 13

TABLEAU I (suite 5)
BREVE DESCRIPTION DES SEQUENCES SELON L'INVENTION

N° de Séquence	Désignation
132	Polypeptide codé par la séquence SEQ ID NO 15
133	Polypeptide codé par la séquence SEQ ID NO 16
134	Polypeptide codé par la séquence SEQ ID NO 18
135	Polypeptide codé par la séquence SEQ ID NO 19
136	Polypeptide codé par la séquence SEQ ID NO 20
137	Polypeptide codé par la séquence SEQ ID NO 21
138	Polypeptide codé par la séquence SEQ ID NO 22
139	Polypeptide codé par la séquence SEQ ID NO 23
140	Polypeptide codé par la séquence SEQ ID NO 24
141	Polypeptide codé par la séquence SEQ ID NO 25
142	Polypeptide codé par la séquence SEQ ID NO 26
143	Polypeptide codé par la séquence SEQ ID NO 27
144	Polypeptide codé par la séquence SEQ ID NO 28
145	Polypeptide codé par la séquence SEQ ID NO 29
146	Polypeptide codé par la séquence SEQ ID NO 30
147	Polypeptide codé par la séquence SEQ ID NO 31
148	Polypeptide codé par la séquence SEQ ID NO 32
149	Polypeptide codé par la séquence SEQ ID NO 33
150	Polypeptide codé par la séquence SEQ ID NO 34
151	Polypeptide codé par la séquence SEQ ID NO 35
152	Polypeptide codé par la séquence SEQ ID NO 36
153	Polypeptide codé par la séquence SEQ ID NO 37
154	Polypeptide codé par la séquence SEQ ID NO 40
155	Polypeptide codé par la séquence SEQ ID NO 47
156	Polypeptide codé par la séquence SEQ ID NO 52

TABLEAU I (suite 6)

BREVE DESCRIPTION DES SEQUENCES SELON L'INVENTION

N° de Séquence	Désignation
157	Polypeptide codé par la séquence SEQ ID NO 53
158	Polypeptide codé par la séquence SEQ ID NO 62
159	Polypeptide codé par la séquence SEQ ID NO 65
160	Polypeptide codé par la séquence SEQ ID NO 74

REFERENCES :

- BUSH et al., 1997, J. Chromatogr., 777: 311-328.
- 5 • FLOTTE et al., 1992, Am. J. ResPIR: (Séquences PIR non redondantes, version 57). Cell Mol. Biol., 7: 349-356.
- HIGGINS et al. 1996, Methods Enzymol. 266: 383-402.
- KOCH Y., 1977, Biochem. Biophys. Res. Commun., 74:488-491.
- KOHLER G. and MILSTEIN C., 1975, Nature, 256: 495.
- 10 • KOZBOR et al., 1983, Hybridoma, 2(1):7-16.
- McLAUGHLIN BA et al., 1996, Am. J. Hum. Genet, 59 : 561-569.
- MERRIFIELD RB, 1965a, Nature, 207(996):522-523.
- REIMANN KA , et al. 1997, AIDS Res Hum Retroviruses. 13(11):933-943.
- 15 • Ausubel et al., 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y
- Beaucage et al., *Tetrahedron Lett* 1981, 22: 1859-1862
- Brown EL, Belagaje R, Ryan MJ, Khorana HG, *Methods Enzymol* 1979;68:109-151
- 20 • Bush et al., 1997, J. Chromatogr., 777 : 311-328.
- Flotte et al., 1992, *Am. J. Respir. Cell Mol. Biol.*, 7 : 349-356.
- Fuller S.A. et al., 1996, *Immunology in Current Protocols in Molecular Biology*, Ausubel et al.
- 25 • Hames BD and Higgins SJ, 1985, "Nucleic acid hybridization : a practical approach", Hames and Higgins Ed., IRL Press, Oxford.

- Higgins *et al.*, 1996, *Methods Enzymol.* **266**:383-402.
- Houbenweyl, 1974, in *Meuthode der Organischen Chemie*, E. Wunsch Ed., Volume 15-I et 15-II,
- Koch Y., 1977, *Biochem. Biophys. Res. Commun.*, **74**:488-491
- 5 • Kohler G. and Milstein C., 1975, *Nature*, **256** : 495.
- Kozbor *et al.*, 1983, *Hybridoma*, **2**(1):7-16.
- Leger OJ, *et al.*, 1997, *Hum Antibodies*, **8**(1): 3-16
- Martineau P, Jones P, Winter G, 1998, *J Mol Biol*, **280**(1):117-127•
- McLaughlin BA *et al.*, 1996, *Am. J. Hum. Genet.*, **59** : 561-569.
- 10 • Merrifield RB, 1965a, *Nature*, **207**(996): 522-523.
- Merrifield RB., 1965b, *Science*, **150**(693): 178-185.
- Narang SA, Hsiung HM, Brousseau R, *Methods Enzymol* 1979;**68**:90-98•
- Reimann KA, *et al.*, 1997, *AIDS Res Hum Retroviruses*. **13**(11): 933-943
- 15 • Ridder R, Schmitz R, Legay F, Gram H, 1995, *Biotechnology (N Y)*, **13**(3):255-260
- Rust S. *et al.*, *Nature Genetics*, vol. 20, Septembre 1998, pages 96-98
- Samulski *et al.*, 1989, *J. Virol.*, **63** : 3822-3828.
- 20 • Sanchez-Pescador R., 1988, *J. Clin. Microbiol.*, **26**(10):1934-1938•
- Urdea M.S., 1988, *Nucleic Acids Research*, **11**: 4937-4957
- Urdea MS *et al.*, 1991, *Nucleic Acids Symp Ser.*, **24**: 197-200.
- Chomczynski, P., & Sacchi, N. (1987) *Anal Biochem* **162**, 156-9.
- DeRisi, J., Iyer, V., & Brown, P. (1997) *Science* **278**, 680-6.

- Francis, G. A., Knopp, R. H., & Oram, J. F. (1995) *Journal of Clinical Investigation* 96, 78-87.
- Iyer, V., Eisen, M., Ross, D., Schuler, G., Moore, T., Lee, J., Trent, J.,
- 5 • Staudt, L., Hudson, J. J., Boguski, M., Lashkari, D., Shalon, D., Botstein,
- , & Brown, P. (1999) *Science* 283, 83-7.
- Jonas, A. (1986) *Reconstitution of High Density Lipoprotein*, Vol. 128,
- 10 Academic Press, London.
- Marcil, M., Yu, L., Krimbou, L., Boucher, B., Oram, J., Cohn, J., & Genest,
- J. J. (1999) *Arterioscler Thromb Vasc Biol* 19, 159-69.
- Remaley, A. T., Schumacher, U. K., Stonik, J. A., Farsi, B. D.,
- 15 Nazih, H., &
- Brewer, H. B. (1997) *Arteriosclerosis Thrombosis and Vascular Biology* 17,
- 1813-1821.
- Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning, a Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- 20 • F.Ausubel et al . Current Protocols in Molecular Biology, J.Wiley and Sons Eds, 1999
- Fromont-Racine et al. Nucleic Acid Res.21(7) :1683-1684 (1993)
- 25 • Current Protocols In Molecular Biology Volume 1 edited by Frederick M. Ausubel, Roger Brent, Robert E. Kingston, David D. Moore, J.G. Seidman, John A. Smith, Kevin Struhl - Massachusetts General Hospital Harvard Medical School, chapitre 11

- Köhler et al, Nature 256 :495 (1975)
- Köhler et al, Eur. J. Immunol. 6 :511 (1976)
- Köhler et al, Eur. J. Immunol. 6:292 (1976)
- Hammeling et al., in : Monoclonal Antibodies and T-Cell Hybridomas,
5 Elsevier, N.Y., pp. 563-681 (1981)
- Wands et al. (Gastroentérolology 80:225-232 (1981)
- Morrison, Science 229 :1202 (1985)
- Oi et al., Biotechnique 4 :214 (1986)
- Cabilly et al., US patent n°4,816,567
- 10 • Taniguchi et al., EP 171496
- Morrison et al., EP 173494
- Neuberger et al., WO 8601533
- Robinson et al., WO 8702671 ;
- Boulianne et al ; Nature 312 :643 (1984)
- 15 • Neuberger et al., Nature 314 : 268 (1985)

REVENDICATIONS

1. Acide nucléique codant pour une protéine ayant une séquence en acides aminés choisie dans le groupe des séquences d'acides aminés
5 SEQ ID N°129 à SEQ ID N°160 ou un fragment peptidique ou un variant de cette dernière, ou un acide nucléique de séquence complémentaire.

2. Acide nucléique comprenant au moins huit nucléotides consécutifs d'un polynucléotide choisi dans le groupe constitué des
10 séquences nucléotidiques SEQ ID N°1 à SEQ ID N°81, ou un acide nucléique de séquence complémentaire.

3. Acide nucléique selon la revendication 2 comprenant au moins vingt nucléotides consécutifs d'un polynucléotide choisi dans le groupe
15 constitué des séquences nucléotidiques SEQ ID N°1 à SEQ ID N°81, ou un acide nucléique de séquence complémentaire.

4. Acide nucléique ayant au moins 80% d'identité en nucléotides avec un acide nucléique selon l'une quelconque des revendications 1 à
20 3, ou un acide nucléique de séquence complémentaire.

5. Acide nucléique hybridant, dans des conditions d'hybridation de forte stringence, avec un acide nucléique selon l'une quelconque des revendications 1 à 4, ou un acide nucléique de séquence
25 complémentaire.

6. Sonde ou amorce nucléotidique comprenant au moins huit nucléotides consécutifs d'un acide nucléique selon l'une quelconque des revendications 1 à 5.

7. Sonde ou amorce nucléotidique selon la revendication 6 ayant une longueur de 15 à 300 nucléotides.

8. Sonde ou amorce nucléotidique selon la revendication 6 ayant une longueur de 20 à 200 nucléotides.

9. Sonde ou amorce nucléotidique comprenant au moins huit nucléotides consécutifs d'un polynucléotide choisi parmi les séquences SEQ ID N°82 à 101 et 102 à 128.

10

10. Procédé pour l'amplification d'un acide nucléique selon l'une quelconque des revendications 1 à 5 contenu dans un échantillon, ladite méthode comprenant les étapes de :

15 a) mise en contact de l'échantillon dans lequel la présence de l'acide nucléique cible est suspectée avec une paire d'amorces nucléotidiques dont la position d'hybridation est localisée respectivement du côté 5' et du côté 3' de la région de l'acide nucléique cible dont l'amplification est recherchée, en présence des réactifs nécessaires à la
20 réaction d'amplification; et

b) détection des acides nucléiques amplifiés.

11. Procédé d'amplification selon la revendication 10, caractérisé en ce que les amorces nucléotidiques sont choisies parmi les amorces selon l'une quelconque des revendications 6 à 9.

12. Nécessaire pour l'amplification d'un acide nucléique selon l'une quelconque des revendications 1 à 5 comprenant :

30

a) un couple d'amorces nucléotidiques dont la position d'hybridation est localisée respectivement du côté 5' et du côté 3' de l'acide nucléique cible dont l'amplification est recherchée;

5 b) le cas échéant, les réactifs nécessaires à la réaction d'amplification.

13. Nécessaire pour l'amplification d'un acide nucléique selon la revendication 12, caractérisé en ce que les amorces nucléotidiques sont
10 choisies dans le groupe constitué des amorces selon l'une des revendications 6 à 9.

14. Sonde nucléotidique selon l'une quelconque des revendications 6 à 9, caractérisée en ce qu'elle comprend un composé
15 marqueur dont la présence est détectable.

15. Procédé de détection de la présence d'un acide nucléique selon l'une quelconque des revendications 1 à 5 dans un échantillon, ladite méthode comprenant les étapes de :

20

a) mettre en contact une ou plusieurs sondes nucléiques selon l'une des revendications 6 à 9 et 14 avec l'échantillon à tester;

b) détecter le complexe éventuellement formé entre la ou les
25 sondes et l'acide nucléique présent dans l'échantillon.

16. Procédé de détection selon la revendication 15, caractérisé en ce que la ou les sondes sont immobilisées sur un support.

17. Nécessaire pour la détection de la présence d'un acide nucléique selon l'une quelconque des revendications 1 à 5 dans un échantillon, ledit nécessaire comprenant :

5 a) une ou plusieurs sondes nucléotidiques selon l'une quelconque des revendications 6 à 9 et 14;

b) le cas échéant, les réactifs nécessaires à la réaction d'hybridation.

10

18. Nécessaire de détection selon la revendication 17, caractérisé en ce que la ou les sondes sont immobilisées sur un support.

19. Vecteur recombinant comprenant un acide nucléique selon
15 l'une quelconque des revendications 1 à 5.

20. Cellule hôte recombinante comprenant un acide nucléique selon l'une quelconque des revendications 1 à 5.

20 21. Cellule hôte recombinante comprenant un vecteur recombinant selon la revendication 18.

22. Polypeptide choisi dans le groupe des polypeptides suivants :

25 a) un polypeptide comprenant une séquence d'au moins 15 acides aminés choisi dans le groupe constitué des peptides de séquences SEQ ID N°129 à SEQ ID N°160, ou un fragment peptidique ou un variant de ce dernier ;

b) un polypeptide ayant au moins 80% d'identité en acides aminés avec un polypeptide tel que défini en a).

30

23. Polypeptide comprenant des modifications d'acides aminés de 1, 2, 3, 4, 5, 10 à 20 substitutions, additions ou délétions d'un acide

aminé par rapport à la séquence en acides aminés d'un polypeptide selon la revendication 21.

24. Anticorps dirigé contre un polypeptide selon l'une des
5 revendications 21 ou 22.

25. Anticorps selon la revendication 23, caractérisé en ce qu'il comprend un composé détectable.

10 26. Procédé pour détecter la présence d'un polypeptide selon l'une des revendications 21 ou 22 dans un échantillon, comprenant les étapes de:

a) mise en contact de l'échantillon avec un anticorps selon l'une
15 des revendications 23 ou 24;

b) détection du complexe antigène/anticorps formé .

27. Nécessaire de diagnostic pour la détection de la présence
20 d'un polypeptide selon l'une des revendications 21 ou 22 dans un échantillon, ledit nécessaire comprenant:

a) un anticorps selon l'une des revendications 23 ou 24;

25 b) un réactif permettant la détection des complexes antigènes/anticorps formés.

28. Procédé pour le criblage d'une molécule ou d'une substance candidate interagissant avec un polypeptide selon l'une des
30 revendications 21 ou 22, ladite méthode comprenant les étapes de :

a) mettre en contact un polypeptide selon l'une des revendications 21 ou 22 avec la substance ou molécule candidate;

b) détecter les complexes éventuellement formés entre ledit polypeptide et ladite substance candidate.

29. Nécessaire pour le criblage d'une molécule ou substance candidate interagissant avec un polypeptide selon l'une des revendications 21 ou 22, ledit nécessaire comprenant :

10

a) un polypeptide selon l'une des revendications 21 ou 22:

b) le cas échéant, des moyens nécessaires à la détection du complexe formé entre ledit polypeptide et la molécule ou substance candidate

15

LISTE DE SEQUENCES

<110> AVENTIS PHARMA S.A.

<120> Produits d'expression de gènes impliqués dans des
maladies génétiquement liées au locus 9q31-34.

<130> Extension FR61610K - AVENTIS

<140>

<141>

<160> 160

<170> PatentIn Ver. 2.1

<210> 1

<211> 552

<212> ADN

<213> Homo sapiens

<400> 1

```

gggtaaatag ttgtgaacat aactattgtg aatactatgt tatatgttat agttgtgaat 60
gtaaatagta ttcacaattg tgaatacaat gtaaatactg ttcacaattg tgaatacaat 120
gtaagtagct atgtaaatag tgaatacaat gtaaatactg atgtaaatag ttgttatgct 180
gattagggaa taatgagaag gaaaaaaagt atacatatc agtacagatg cagtcattct 240
tttttccct cgaacatttt tgatccacaa ttggttgaat ccacagatgg ggaacacaca 300
gtcttggtaa atttaaccaa caaggagggt aaacgcattc caacagggaa ggtaaactgc 360
acatccatca gtacctcttn gaggggcatt actggtttat aggcttcaat tacagtggac 420
tatntccag gantaatggg ccttttagtt tttcccnga angnttaant tggnggnttg 480
ccttttgttt ggantgagnt tantcnaacn tattgangta atttttaaag gggntttcat 540
aacagaaggc ct                                     552

```

<210> 2

<211> 1246

<212> ADN

<213> Homo sapiens

<400> 2

```

tagtggaatc agtaaggctt ggcagttgac cttgtttgtt ggagagaagg gataagattt 60
taaagctaca tgtctgaaag aatgatgctg ctgattgaaa taaaggaaga aaggatgcat 120
ttcgggctcc aacctgtcct aggaaggcct agacctcaaa caccaacacc tccatgcatt 180
tcctctttgg ctactatgac ttttccctga cttctgcctc tccagctctc tgggctgctg 240
cttccacctg ttcattctgac ttagaccctc cctgctgggt ccttggtcac ctactcattt 300
gggtgcttcgt ctgccatcag tacctccatt gcagctgggt ggatgtcagt caccatctct 360
tatatttgct tccactaga aagatcaaga gaagttattt ctttcccttg cgctccaatt 420
tttctctaga cagttggtat ccacaatttt aaaaaatggt ccatgttgtt taaacaagca 480

```

```

ttcgtgaga ggggctgtta atacacatcg tgccectttt ataaaaattc atgcatggaa 540
tcctacatta ttatgcatca aaatctccag aaatgtctta ggatttttgc agggagaata 600
ttaaatacat tgttttgctt tgttttgaag agactagatg tgcagaggaa gagagggtgc 660
atggtgggag ggtacatttg agttgtcaac agtctctgca gtgtcaggtc aattacatca 720
gcacttggac tggaccaggg aaaggaatga ttctgttcc tgggaatgtc agaaggacct 780
gatgattata tttggcaaag ccaggaggag tggctttgaa tgtcattgct aagaattaca 840
ctttgagtag catttctgga tgtctgagct tttcaaatga tacttctttt ctgctgtggc 900
tttcctttct gttggactgg ttcccagagg gtccctctgt ttgtccttgc cctcgctttt 960
atatcagttc atgttttctc ttctgtcatc ttccctccca gcgtgtttc tccacccct 1020
cctgctgcac tcacaacagc ttccctctc ctgtttagag gtggaagcat gtaagaatgc 1080
gtttgagggg gatgcttgcc aaaggacagc atattcaaca tctggtatca acaaggtaat 1140
gtttaacctt agactagcca aactagtgat gacctgttc catgctgcat ctgctgcttt 1200
ttgtgtgat gggactcaga aatcatgaga aaggtcttca gtgate. 1246

```

<210> 3

<211> 3035

<212> ADN

<213> Homo sapiens

<400> 3

```

aactgagcaa agattgcctt tgcttcacct attgacacat ttaagaacct .cgttttatag 60
agagtacgga gtgggtctg agattctaca tttcttaca gttcctggtt tatgccagtt 120
ctgtttgtct ggtcaccaca cttctgaata gcaagggttc agaggactct atgaagtttt 180
tatttctcaa aaatttggtt taggtacatc ttacatata gtggtgctag atattattct 240
ttcagaactt aaccatgcaa caatgcagaa gaatacaaaa aatgcttgca cacagtagtg 300
tattttaatg gaactataaa agcattgaac tgattcctaa cttacatat aaaggataat 360
taaaccagca gcctaaagca ccttgctcta ctgttaatgg tacttatgag ttataacag 420
tagcttccaa ccttggtctat gcattagaaa cagcatttta taaatgatta gtcccactgc 480
gtcccatttc cacaggatct tgattcaatt tacctgggct ggggccttgt cctctatatt 540
tttcagaagc ttacagggg attatgatgc ataattggga ttgagaatna ctgatgtagc 600
attcacaact gtgtctcatt ttattcttaa tcccatgagg ccatgcagaa ggaaggaaat 660
gatgctcaga gaagaaccac ctggtgtcaa gcagatagct agccgaatgg caaagcagaa 720
cctcaatccc agatttcttg actccaaaat atgtagtcca aactgcctgc ccttggtcca 780
ttccttgagc tctatcatga attgggttta tctgagagt cagatcacat tacagggtac 840
ttgtttcagt gtctccattt gtgagcaagt atatcacaat tcatttcaact tttgagcagt 900
ttcttacccg aggtagaaat gcaataacca tttcttggtg ggtggtgtag gggatagaca 960
ggtaaataat aagacacagt tcatcatccc ccacctggc aaattacgga cacagcccag 1020
tagggagtat aagggttgta cacagctatc atgcgcttga tcagtgcatt tcttcaaagt 1080
gcacttgtag ctcagcagag agctaagaag catgaaggaa gtcttcatgg aggagatggc 1140
actggagcta agtcttcaga gtcggcattt tggcagggtg agattgggaa accagaggaa 1200
cttcactaga tgcaagggtc ttgtgaagat tcaaggtaat gagacacata attgagaggg 1260
agtgcagca tggggaattt ctgagtttg gcagtaatgt gatcagaact gtactccagg 1320
aagaagagta gaagaaagaa ctgagtggag agaagactgg agtaggggat acgtaggagg 1380
gtttgtccct ggtaggatg agatgatagc aggaccaga agaggttgca gtaaaaccaa 1440
gagagacagc cgggtgcggg ggctcacgcc tgtaatccca gcacattggg aggccaaaggc 1500
gggtggatca cctgaggtca ggagttcgcg agcagcctgg ccaacatggt gaaaccccg 1560
ctatactaaa aatacaaaaa gatagccaga cgtggtggca ggcgcctgta atcctagcta 1620

```

```

cttgggagggc tgaggcaaga gaattgcttg aaccaggag gtggacgttg tagtgagcca 1680
agatcacacc attgcactcc agcctgggta acaagagtga aactccgtct ccaaacaac 1740
aaacaaacaa acaaaccac gagggagatc agtgtgaagg gcattgccaa agtggaaatca 1800
gtaaggcttg gcagttgacc ttgtttgttg gagagaaggg ataagatttt aaagctacat 1860
gtctgaaaga atgatgctgc tgattgaaat aaaggaagaa aggatgcatt tcgggctcca 1920
acctgtccta ggaaggccta gacctcaaac accaacacct ccatgcattt cctctttggc 1980
tactatgtct ttccctgac ttctgcctct ccagctctct gggctgctgc ttccacctgt 2040
tcacttgact tagacctcc ctgctgggtc cttgttcacc tactcatttg gtgcttcgtc 2100
tgccatcagt acctccattg cagctgggtg gatgtcagtc accatctctt atatttgctt 2160
cccactagaa agatcaagag aagttatttc ttcccttgc gctccaattt ttctctagac 2220
agtgtgtatc cacaatttta aaaaatgttc catgttgtat aaacaagcat tcgctgagag 2280
gggctgttaa tacacatcgt gcccctttta taaaaattca tgcattggaat cctacattat 2340
tatgcatcaa aatctccaga aatgtcttag gatttttgca gggagaatat taaatgcatt 2400
gttttgcttt gttttgaaga gactagatgt gcagagggaag agagggtggca tgggtgggagg 2460
gtacatttga gttgtcaaca gtctctgcag tgtcagggtca attacatcag cacttggaat 2520
ggaccagggg aaggaatgat tctgcttctt gggaaatgtca gaaggacctg atgattatat 2580
ttggcaaagc caggaggagt ggctttgaat gtcattgcta agaattacac tttgagtagc 2640
atctctggat gtctgagctt ttcaaagat actctcttct tgctgtggct ttctctctct 2700
ttggactggg tcccagaggg tctcttctgt tgccttggc ctcgctttta tatcagttca 2760
tgttttctct tctgtcatct tcttcccag cgctgtttct ccacccctc ctgctgcact 2820
cacaacagct tcccctctcc tgttttagag tggaagcatg taagaatgag tttgaggggg 2880
atgcttgcca aaggacagca tattcaacat ctggtatcaa caaggtaatg .tttaacctta 2940
gactagccaa actagtgatg acctgcttcc atgctgcac tgctgctttt tgtgttgatg 3000
ggactcagaa atcatgagaa aggtcttcag tgatc 3035

```

<210> 4

<211> 1479

<212> ADN

<213> Homo sapiens

<400> 4

```

gtctgcagaa ttgcgcttg ctgccacct tgtaccagct attccagatg accaagccct 60
gctagaataa agactgactt ttcatgtga cctctgaagg atgtgcagct cagctttcac 120
ctatccccac gttcacccaa gccaaaaaga agaattgttc attcaaactg gaaatgatct 180
ccagcttgga aatgagactc ctcatgagct tattctaaga tcagagtctg aaaattattt 240
tctcaaaagt tacaaataaa tgtcttttaa cagtactgtc agactagtga ccaacaagtt 300
actttatgag gatggatgat gatattagag ttcaagcaat tgtccagttt gtttcgatgc 360
cacagtttgc ttttaaattc ccatagttag cttccatctt gagacccact caacacagaa 420
ggattacata aaggttgatc acccacgctc tacagatgaa caaattgaga ccagagatg 480
ctaaactgct ctgcaggtca caaacctgaa cagcagcata gccaaaatgg gaacggagct 540
cttctgactt caagccctg cacttttttc ttaaccacat taattttatg aatatctatg 600
tgggtcaaagt ctcttttgta aaatctgagt ataagctgaa ctagatcatg tctctgtaag 660
ctaagatggc agaaacacat tgtcatgtta gtttgaggga gcaaacagat ataagtatgt 720
gttgtaatgt tatagttagc tggtatgtat ccccgagcaa tgtccaaagt gatcagagct 780
actagatgac ttattacaat gttcagatcc cactccttca ttccctctct ctcagtgaac 840
tgtctacccc caagctttag aactgagctc tctccaactt agaactcatt agagtttgtt 900
gcattagtga ttggatctcc tcatgaaagt tttttgacca atgacagtag agccctttta 960

```

```

agctageccag gcccaagata gggaagcact gctgggtttac atttctgctc tttctaattc 1020
agttgtctat tgcattgcaga catccctcgc ccgacttctc cattatggaa ggctaaggcc 1080
atccttccctt gtccagtttc tctacctaca cactaacatg ccttttcccc atggagtgaa 1140
ctctgccagt tgtcttttga cattttctgc agatgcaggc tcagttctgc tcttttggtc 1200
ccaccacttt ctagatcttc ctccatttaa tcccttattc ttataaggag tctaattttt 1260
gtgtattgtc cttgattagg gcagctgaca gctttacacc aaagctgaag tctctcactt 1320
aatcctaagc aaggggctag ggagaactaa gatatccttc ccatatcaag ggatataggt 1380
tactagaagg gtggaggtct cttaacaggc tcactcacct ttcagagttt aaattatggg 1440
caggcttaag aagctctctc ttgtgactga gtgtattgc 1479

```

<210> 5

<211> 5169

<212> ADN

<213> Homo sapiens

<400> 5

```

cagttctttt cattccatca ctttaggtga tgggtaagat ttttgaaagc cttatatttt 60
ttgattttgt tgtctagttt aatccctacct ttaatagttg tgtttggtaa aattccact 120
tgaatgtgac actgataata attatgctga ttttttagcat ctcttatagg aatcaaagtt 180
tattaaagtt acatagagga ttgaaaaatg tatatcactc aatttttattc taagaaggat 240
aggttataaa gggagggtacc taaatactca aataatgtat atattctttt tcataacata 300
tggaatgctt taagcaattg ttttgaaaaa aatctgcgta tctttgactt aattggcaca 360
tgactttttc aagcagccat tcattcagag gtttggtttc tctcagtcct tttgctacat 420
tcactttctt ttcaaagaaa tgtaattat taaaaaatt gacatagata tctttcccaa 480
acttggggta aaacccatgg ttatgtggaa cataactgtt cttaaaaagt caattataat 540
ttgtaactca catcctttga gctaaactaa attaaaaatta cagtatttaa taattctttg 600
ggcattttta agagggtctga atttgtattt ttcttcattt ttgaaacttt acatttggtg 660
actttttttc attctcattt aaaatatatt gtgctatgat aaccaacctt cttccaagga 720
gtgaccatta ctgcctacat ttgcgttgct ttctacatag agacttgat aatagtatta 780
atagtagctc attctctcta aaaatttact gccactgag aggatgtttt gtttccttga 840
taaaacaaat gttaatgagg aattttattt gaattggaat attgtttttc tagaggacat 900
tcatatctgc actattatct gatgacatgt tggtaatttt aaagactgca aggcagttta 960
gagaatgaag taaaccaata gatactcttt tgattctcca aagaaatata attttggttt 1020
ttgttctca gagagagttt gaaagaaatt ttcagatgtt ctgttcccta tagagggcct 1080
attccaggat ctaaatgaat gggaatttat aattcctttc atagactcaa acctgaaagc 1140
agaaattttt agtagttgag ttgctttaag tgaattttta caaatatgac acagaaaagg 1200
tcagatgctg ttgtataaaa ttttttacta gtgtgtctta tataattctt tcttctcatt 1260
aattcagtg ctttttcttc taagcatact cttgggtccat gccccatgca gtgtcaactg 1320
atgtttaagc tacagagcat tgttgtagtg gtgagggcc tctgttgagc gggcatgggg 1380
aatgcatat attaatcagg ggctggagga gtatacatga ttcttggtgt catgttggtc 1440
tttgaaactc ttcacataaa attttctaatt attcatttgg aaaacctctc cactggagag 1500
gtttcatctc tgagagggtt tgtaataata gtgtaagtgc agtaaatcca gtccaaattc 1560
tcatcacatg tattatttga tactaaattt tcaattatta cttcaaaata agagtctgag 1620
gatttcttct tactggcggt cttaaagggt tctaaaatta aagaacaaga tgtggttttt 1680
tgtttaagac gtttttagtt tatttggtgt taagttaaagc tggagaaagt ttgtcatcct 1740
cattttcaaa ctgaaaacta atcaaatata gtgcctattt agtctcaaaa taaggtaact 1800
attaacttga ttatactata taatattctc aattacattt gaatttaaaa atattgagct 1860

```

```

cttgattaca atttaataca ttgtaaaaaag tattaataccc tttgaaacat cttttgttgc 1920
ttatttagtt tttgtttcta cagggtcaaac atgattgctt tctataaaaag aaatgcttga 1980
gaggttgact ataattggata tgtcacaggt ataaaaacagt tgttttctaa aaacatgcat 2040
ttagtatggc atttctctttt agcaactgaa ttcccaacga gttttattaa gctggatatac 2100
gaaatgaagg agctgcctta agcacttttag aaaaagaatt tttttacaat tcattttgac 2160
tctcatggct gaccatgtca ttatgtcttt aaattttggt aaatatgtag ataccaagca 2220
ttaataacta atgcacacag aaatttaata tacagcaatt ctttgaatgt tccagggtgta 2280
cgtaaaactaa ctgaaagtat taaggacatg cctgtctata catgggtttc attgatggta 2340
tctgtatcat cttgacaaat cgtgaactgc tgcgtgatgg ctagactttg cctatttact 2400
ctgttatgca aacagtaatt tttccctatg ttatgaagag agacatatgt gcttgcttta 2460
atttacttat ttatgaaatt aaagcatgaa taatatgtat aatttgaatt ttttgtgact 2520
gatecctggc agtataaaatg ataaaagtaa atgtaatgga atcttttaat taggctaaga 2580
tatgctattt cataacttat gtagaatgat ttttatctat atactttgtc cataaaaatta 2640
taattgttta tgaattttta ttgtttctta tggaaaacaat tggaaaagta tatggaaaat 2700
gattatttca aagaatattt atttaaagag ggaagtgtag acttcttact gtaaaatatg 2760
tgattccagg ctttaagaaat tgattatatt ttaaaataat tattttttta aaaacgtaat 2820
ttgtttttta aagatgagtc tccattcatt ggctaatacca tacagcaaat atttgagtgt 2880
ttgtgatatg gcagtaaaaa tgtgattact ttgtatgtac attaaccaaa accagtgaag 2940
tttgatggac tgtgcataga gttttgcagc ttttgagtag tgtgacagag tttgcctcct 3000
aaaattgttt tgcataaag tcagaagttt aaaatgtgtt atagtttgct aaaaaaaaaa 3060
caaattaact gcatataaaa tggttcactt ttatatagtc acatgtaatg ggctgaaaaa 3120
agtttcttct ttgattatgt agttactcta gatacattct tttagggag tgatgggtca 3180
acatatattt actcagtata cagataacca caaagttgtt atttttaacc aaaatataca 3240
tttcataatg ttggcactgg agtattataa gagatcatga atgtttcctt aattatcatg 3300
gcaagattac ctcacgttca gtatagttta gttttgtgtg tatcttacag tatgcatgtt 3360
ctgaaattat ttgtgatcct gatattgtcac atataatagc tttgttacct ggagtatttg 3420
attaacctaa acagtttaag cattttgaat tagttggagt ttaaattggat tacatttggt 3480
gtatatgttc ttcattgaaa agatattgtt ggatatgccat catcagagtt agcattgact 3540
tataaataaa taccagaatt taccagagta cacttttctc ttcagatgac tcaattatta 3600
aataaaaagaa tatttcttat ttctgggtca atataaggta ctgacttctg atgcataatc 3660
acttaagcta tgtgagttta aactgggtatc ttttcttagt ggtaccatt caatgtaagc 3720
tgggtccatg gaatggacat gaagaggatt tcacagtatg tagagcagaa ggcacgtgaa 3780
tgtgtttgct ttggcttggga gcttattaag ttttgactac ggtaaaaata agtcaaatag 3840
taagtggtaa aacacatttt tgttagtatt ggaactttct ggagaacata agggctatga 3900
gaatgcatat atatatTTTT taacatttcc tatatatcta aggtaccaa gcactgagtc 3960
taatttacct attaaggag actctttaa atcaacttta taactaatc atactataag 4020
acagataata gctaaagttt tggataaatt tatattaaaa gccgcaagtc ttaaaaaatcc 4080
ctggatatga cataaaaagg attttggctt cttttttgaa gtatttaaaa ttaatcacct 4140
tagctctacc atataactaga tctgtgaccg ctacacaaat tgtttatcat ctttgggtct 4200
ctattgcctt ctttataaaa taagtgtgaa ttgttcagcc tgcctcacag ggctgtgtg 4260
aggaaataaa atgaaatgga gtatgtgaaa ttgcttcagc aacagcaaag tgctacgtaa 4320
atgtaaagtg ttgttttttag ctaataatgg atttaagtgt ttggataatt gtagatgcat 4380
ttactttgat aaagcgtgtg cttaaagtgg tatcaccagt gatttctaac atgattttta 4440
aaaaataaaa ccccaattaa aatgttcttt aatattcatt taatttgtgc atgacttgtg 4500
gcctttttgt attttctcaa gcctattact ctagagctgt aaaagctctt gcacagcatt 4560
gttgtgtcag tgtagaatag tggcataaat aaatgaatta agctacattt ccatgcaggc 4620
atttgcactt tgaagttagg gctgcccagt gccacatgca aagagtgtat ttggtctaaa 4680
cagttctttc caaacaacca aatggctggg cagcttgttg tcacggataa atgatataaa 4740

```

```

atgtaaccac tgaacatttg acaactttgt tgcattatca tttttaaaaa aagatgaagc 4800
atttcaaaat cttgttaacc tgggtatctc ttggtatctc tcctgacata tttgttttgc 4860
tttaaattgg attaaatggg ttacagggtta tctgaatcta cagatttagt atgtctatta 4920
tctggacatg attttgctat gcagttgtga taataaaaaa tctaattcct caggtttggc 4980
cttttaagtt atggctgaag acctttaatg atacttatga tgcactcagt gccatacaat 5040
agtgttttaa gaagcacatg gcctcatttt ccttagaaac aagacaaagc ctttttaaaa 5100
tgtttttttt tcactttaaa ataggggtatt caaagtggag taacgtgtat tttgaaatga 5160
ttttgtgtt                                     5169

```

<210> 6

<211> 7723

<212> ADN

<213> Homo sapiens

<400> 6

```

gtggggccgg gcggtgccgg tgcgggctgg ggcagtgagc tgagtagcgg tcttgggggtg 60
tgcgatctcg ctgagcctcc tcacacgggt cgctcgctcg ggttcgagcc cagtgggctt 120
agccactcgc catggactcc cagaaatact gctttaagga gaatgaaaac gttacagttg 180
ataaagcctg ttttctgac tcatacatca ctattggacc agagtctatt aacttgcagc 240
aggaagctct acagaggatc atttcaactc tggcaaataa aaatgatgaa attcagaact 300
ttattgatac actacatcat aactaaaaag gagtccagga aaattcgtcc aacatactct 360
cagagttaga tgaagaattt gatagtttat actctatact ggatgaagta aaagaaagta 420
tgattaaactg tatcaagcag gaacaagctc gtaaatccca agagttacag agtcagatta 480
gtcaatgtaa taatgcctcg gagaactctg aagaactatt agaatttgca acaagggtcat 540
tagatataaa ggaacctgaa gaattttcaa aggctgccag acagatcaag gatagagtca 600
caatggcttc agccttttgc ctttctttga aaccaaaggc cagtgcacac atgactcatt 660
taatgggtgga tttctcacag gaaagacaga tgctgcaaac tttgaagttt ttgccagtcc 720
ccaaagctcc agagatagat ccagtagagt gtttggtggc agataactct gtaacagtgg 780
cttgagagaat gccagaagaa gataataaga ttgaccattt tatactggaa cataggaaga 840
ctaattttga tggacttcca cgtgtaaaag atgagcgatg ctgggagata attgataata 900
ttaaggggtac tgaatataca ctatcaggct taaaatttga ttcaaagtat atgaatttca 960
gagtgcgagc ttgtaacaag gctgtggctg gagagtattc tgatccagtg actctagaga 1020
ccaaagcact taacttcaat ttggataact cctcatccca tttgaacctg aaagttgaag 1080
atacatgtgt agagtgggat cctactggag gaaaagggtc agaaagttaa attaaaggaa 1140
aagagaacaa gggcagtgct catgttactt cactgaagaa acatacaaga agtgggtacac 1200
catcccaaaa acgaacatct gtaggctcca ggccaccagc agtaagaggc agtagagatc 1260
gttttactgg agaatacat acagtgtctg gagacactgc tattgaaagt ggacaacatt 1320
attgggagggt caaggcccag aaggattgta aatcctacag tgtgggagta gcatacaaaa 1380
cgttgggggaa atttgaccaa ttgggaaaga caaacactag ttggtgtatc catgtcaaca 1440
actggctaca aaacacattt gcagcaaagc ataataataa agtcaaagct ttggatgtta 1500
ctgttctgaa aaaatagggtg tattttgtga ttttgatggg ggtcaacttt cattctatga 1560
tgcaaattct aaacagttgc tatattcctt taagacaaaa tttactcagc cagtactacc 1620
tggtttcatg gtatgggtgt gtggactttc tttgagtact gggatgcagg ttccaagtgc 1680
tgtgagaaca cttcagaaaa gtgaaaatgg aatgactggg tcagctagca gcctgaacaa 1740
tgttgttact caatagtgtc tactcagaat acgtttaccc tccgtcttga ttaggtggcc 1800
ttttctgtgc agttactaat cacaggaatt tggtagtagt gaaaatcagg tttgctgtgt 1860
tctgctttga ggcctggaat cttttatcat taaacaccta gtacgaagca tttgcaggaa 1920

```

```

cctactgtgc agtatcatag aggcaagcag ataccaagca aaaaactgat gattgaagag 1980
taaattggggg aaaaggcagt gtttaattaa cataaaaact catttttgta tttcttggat 2040
tactttgact attctaattgt ttaattacat atggtaaccc taaggcctgg ggagaaaagc 2100
ttttaaatct tgccttcttt cttctgtact ttgtcttttt aaaatctcat tactatctat 2160
tatttagttc taacacagaa gctttaaaaa tacatagtca tccaagggtt tctaaaaatt 2220
gaaatcatat gttgggaatg gtaaaaagggt tttcaaattgg tttatttttc ctcttttata 2280
aataagtttt acaaaatttt cctcttttgt ttattgacta gattgtatat aattttcttt 2340
tttccaaaat attgtgataa gaaatttcta gacacaacag cttaaaatca cccaaatttc 2400
agttctttac ctaactgcac taacaatggc aaggggggta ttctttatat gttgccttgt 2460
ttaactacag ttcttttcat tccatcactt taggtgatgg gtaagatttt tgaaagcctt 2520
atatttttg attttgtgtg ctagtttaatt cctaccttta atagttgtgt ttggtaaaat 2580
tcccacttga atgtgacact gataataatt atgctgattt ttagcatctc ttataggaat 2640
caaagtttat taaagttaca tagaggattg aaaaatgtat atcactcaat ttttatctaa 2700
gaaggatagg ttataaagggt aggtacctaa atactcaaat aatgtatata ttctttttca 2760
taacatatgg aatgctttta gcaattgttt tgaaaaaaat ctgcttatct ttgacttaat 2820
tggcacatga ctttttcaag cagccattca ttcagagggt tgttttctct cagtcccttt 2880
gctacattca ctttcttttc aaagaaatgt taattattac aaaaattgac atagatatct 2940
ttcccaaact tggggtaaaa cccatgggta tgtggaacat aactgttctt aaaaagtcaa 3000
ttataatttg taactcacat cctttgagct aaactaaatt aaaattacag tatttaataa 3060
ttctttgggc atttttaaga ggtctgaatt tgtatttttc ttcatttttg aaactttaca 3120
tttgttgact ttttttcatt ctcatttaaa atatatgtg ctatgataac caacctctct 3180
ccaaggagtg accattactg cctacatttg cgttgctttc tacatagaga ctgtgataat 3240
agtattaata gtagctcatt ctctctaaaa atttactgcc cactgagagg atgttttgtt 3300
tccttgataa aacaaatgta aatgaggaat tttatttgaa ttggaatatt gtttttctag 3360
aggacattca tatctgcact attatctgat gacatgttgg taatttttaa gactgcaagg 3420
cagttttagag aatgaagtaa accaatagat actcttttga ttctccaaag aaatataat 3480
ttggtttttg ttccctcagag agagtttgaa agaaaatttc agatgttctg ttccctatag 3540
agggcctatt ccaggatcta aatgaatggg aatttataat tcctttcata gactcaaacc 3600
tgaaagcaga aatttttagt agttgagttg ctttaagtga attttaacaa atatgacaca 3660
gaaaagggtc gatgctgttg tataaaattt ttactagtg tgtcttatat aattctttct 3720
tctcattaat tcagtgcctt tttctcttaa gcatactctt ggtccatgcc ccatgcagtg 3780
tcaactgatg ttaagctac agagcattgt tgtagtgggt agggccctct gttgcagggg 3840
catggggaaa tgcataatatt aatcaggggc tggaggagta tacatgatc ttgtgttcat 3900
gttgtcattt gaaactcttc acataaaatt ttctaataat catttggaat acctctccac 3960
tggagagggt tcatctctga gaggttttgt aataatagtg taagttcagt aaatccagtc 4020
caaattctca tcacatgtat tatttgatac taaattttca attattactt caaaataaga 4080
gtctgaggat ttcttcttac tggcgttctt aaagggtctt aaaattaaag aacaagatgt 4140
ggttttttgt ttaagacgtt tttagtttat ttgttggtta gttaaactgg agaaagtttg 4200
tcatectcat tttcaaactg aaaactaatc aaatatagtg cctatttagt ctcaaaaata 4260
ggtaactatt aacttgatta tactatataa tattctcaat tacatttgaa tttaaaaata 4320
ttgagctctt gattacaatt taatacattg taaaagatg taaatccttt gaaacatctt 4380
ttgttgetta tttagttttt gtttctacag gtcaaacatg attgctttct ataaaagaaa 4440
tgcttgagag gttgactata atggatatgt cacagggtata aaacagttgt tttctaaaaa 4500
catgcattta gtatggcatt ctcttttagc aactgaattc ccaacgagtt ttattaagct 4560
ggatatcgaa atgaaggagc tgccttaagc actttagaaa aagaattttt ttacaattca 4620
ttttgactct catggctgac catgtcatta tgtctttaaa ttttggtaaa tatgtagata 4680
ccaagcatta ataactaatg cacacagaaa tttaatatat agcaattctt tgaatgttcc 4740
agggtgtacgt aaactaactg aaagtattaa ggacatgcct gtctatacat gggtttctat 4800

```


gatggtatct gtatcatctt gacaaatcgt gaactgctgc tgtatggcta gactttgcct 4860
atttactctg ttatgcaaac agtaattttt ccctatgtta tgaagagaga catattggct 4920
tgctttaatt tacttattta tgaaattaaa gcatgaataa tatgtataat ttgaattttt 4980
tgtgactgat ccttggcagt ataaatgata aaagtaaag taatggaatc ttttaattag 5040
gctaagatat gctatttcat aacttatgta gaatgatttt tatctatata ctttgtccat 5100
aaaattataa ttgtttatgt aattttattg tttcttatgg aaacaattgg aaaagtatat 5160
ggaaaatgat ttttcaaag aatatttatt taaagaggga agtgtagact tcttactgta 5220
aaatatgtga ttccaggctt aagaaattga ttatatttta aaataattat ttttttaaaa 5280
acgtaatttg tttttaaag atgagctccc attcattggc taatccatac agcaaattat 5340
tgagtgtttg tgatatggca gtaaaaatgt gattactttg tatgtacatt aacccaaacc 5400
agtgaagttt gatggactgt gcatagagtt ttgcagcttt tgagttagtg gacagagttt 5460
gcctcttaaa attgttttgc taataagtca gaagttttaa atgtgttata gtttgctaaa 5520
aaaaaatcaa attaactgca tataaaatgg ttcactttta tatagtcaca tgtaattggg 5580
ctgaaaaagt ttcttctttg attatgtagt tactctagat acattctttt aggggaagtg 5640
atggtcaaca tatatttact cagtatacag ataaccacaa agttgttatt ttttaaccaa 5700
atatacattt cataatgttg gcactggagt attataagag atcatgaatg tttccttaat 5760
tatcatggca agattacctc acgttcagta tagtttagtt ttgtgtgtat cttacagtat 5820
gcatgttctg aaattatttg tgatcctgat atgtcacata taatagcttt gttacctgga 5880
gtatttgatt aacctaaaca gtttaagcat tttgaattag ttggagttta aatggattac 5940
atgtgtgta tattgtcttc attgaaaaga tattgttggg tatgccacat cagagtttagc 6000
attgacttat aaataaatac cagaattttac cagagtacac ttttctcttc agatgactca 6060
attattaaat aaaagaatat ttcttatttc tgggtcaata taaggtagtg acttctgatg 6120
cataatcact taagctatgt gagtttaaac tggtagcttt tcttagtggt acccattcaa 6180
tgtaagctgg tccatgggaa tggacatgaa gaggatttca cagtatgtag agcagaaggc 6240
acgtgaatgt gtttgctttg gcttggagct tattaagttt tgactacggt taaaataagt 6300
caaatagtaa gtggtaaaac acatttttgt tagttatgga actttctgga gaacataagg 6360
gctatgagaa tgcatatata ttttttttaa catttctat atatctaagg taccaaagca 6420
ctgagtctaa ttacctatt aaggagact ctttaaaatc aactttataa ctaattcata 6480
ctataagaca gataatagct aaagttttgg aataatttat attaaaagcc gcaagtctta 6540
aaaatccctg gatatgacat aaaaaggatt ttggcttctt ttttgaagta tttaaaatta 6600
atcaccttag ctctaccata tactagatct gtgaccgcta cacaaattgt ttatcatctt 6660
tggtgtctta ttgccttctt tataaaataa gtgtaagttg ttcagcctgc ctacagggc 6720
tggtgtgagg aaataaaaatg aaatggagta tgtgaaattg cttcagcaac agcaaagtgc 6780
tacgtaaatg taaagtgttg ttttttagcta ataattggatt taagtgtttg gataattgta 6840
gatgcattta ctttgataaa gcgtgtgctt aaagtggat caccagtgat ttctaactatg 6900
attttaaaaa aataaaacce caattaaaat gttctttaat attcatttaa tttgtgcatg 6960
acttgtggcc tttttgtatt ttctcaagcc tattactcta gagctgtaa agctcttgca 7020
cagcattgtt gtgtcagtg agaatagtg cataaataaa tgaattaagc tacatttcca 7080
tgaggcatt tgcactttga agttagggt gccagtgcc acatgcaaag agtgattttg 7140
gtctaaacag ttctttccaa acaaccaaat ggctgggcag cttgttgtca cggataaatg 7200
atataaaatg taaccactga acatttgaca actttgttg attatcattt ttaaaaaaag 7260
atgaagcatt tcaaaatctt gttaacctgg ttatctcttg gtatctctcc tgacatatgt 7320
gttttgcttt aaattggatt aaatggttta caggttatct gaatctacag atttagtatg 7380
tctattatct ggacatgatt ttgctatgca gttgtgataa taaaaattct aattcctcag 7440
gtttggcctt ttaagttatg gctgaagacc tttaatgata cttatgatgc actcagtgc 7500
atacaatagt gttttaagaa gcacatggcc tcattttcct tagaaacaag acaaagcctt 7560
tttaaaatgt ttttttttca ctttaaaata gggatttcaa agtggagtaa cgtgtatttt 7620
gaaatgattt tgtgttctat atagaatgtc ccagttaaga ttttacagaa gcacctggaa 7680

tatctcatta aaaattttaa aactctttaa aaaaaaaaaa aaa

7723

<210> 7

<211> 1046

<212> ADN

<213> Homo sapiens

<400> 7

```

ccttcagtgc tcagaggagc agacacagtt gctcactctc cactcaaagt tgggccaggc 60
cttggtgctaa ttaaaggagg aagcactgct gtagcgcaga aggaaaaggc cagtcctcca 120
cctgttcagc atgataatgc agctcccctg tggtaagggt tggtcggact acaatatctt 180
tgtgactttg atgaaagcat ctgcagttat aatgtggcaa agaaaaggat aattccagaa 240
tttcccattt ccatcatttg agcttgaaat atctgaatag atatttacia aaccttggtga 300
agaagtcacc ttattttctca tgttttctaata gaacatatat aaactttctca tttccagctc 360
taggataagc aaacagcaaa aaattaccaa tgaaaagtta ggagttcaaa agaagtagtc 420
tgaactagat agattagatg tcatcctgtg aaatagggtt accacatgat agggaagaac 480
tttcaaacag tacagggtct attttcagtt attcactatg aaatgagaac atttcacagt 540
tgttaaagggt tgcttttcaga agaaaagccc aactgatgaa tgttttttaa aatgcactat 600
tgagagcgaa caacagactg gatgccataa ttttatagtc aagtttgagg caagttttta 660
gaaactgtgt tacaattcag agaaaagaca gttggtgagc atattaagag atatagggaa 720
cagattttaga tgatccaata tgcttattat cagaattcca gaaaaaggca tgaaatcata 780
aagaagacag aataccaggc aaaattaatg aaaaaatatt aatccaaaca catataataa 840
aaagtatcaa tttcaagaat aaaagaatta tataagcatt taacacacat aaacattggg 900
tacctgtaaa aggtgaaaaa aaaaaacagg ttcagatttt ctgcttaaat atcaaagggc 960
aaatggaaca tttttacgat cacctaaaat tcttgaaata ggccaagatt tttcatatac 1020
aaaggcaaca gccttcttgc cacaac

```

1046

<210> 8

<211> 280

<212> ADN

<213> Homo sapiens

<400> 8

```

gggagaggag gcagaggcca gatcctgaaa aactcttcag tgggtggactg aggtgcctgg 60
acttcagagg cgacaggagg gtagtgaggt ttctggagca ggctgggtcaa agtcagcagc 120
acccttggtt gatatccatc ccattccctc accccacatc tacctgggac ttgcttctct 180
gaccctcctg ctgtggcaag atcctgattg tcccagggtc acatgggtct acttctctgac 240
tgctcatcct cagcgagtc tcagccctct atgaccccc

```

280

<210> 9

<211> 479

<212> ADN

<213> Homo sapiens

<400> 9

```

aatcaagaa catcattctc tcggtatctt tatggagcaa aatgtttatg gtagagacag 60
gtgaccttgc agctagccta accacattct ccaagtgtta attagatggg gcaatattca 120
catagaccta gagtccactg actatgaaag gctctgacct ttttgcggtt tctttaccag 180
gagtgtataa tcttcctagt tcctaaatct ttctaagaga acctgactac ctccaggctt 240
agagggagca gccccttcac tatataacca atttcctga ctgcacatat ggtaattcca 300
tctatggtga ataataaagc cattctaggc caaatagata ctcttatgca gagcccagct 360
gggacagtca cgtgcaagga gaaaagagaa agccagactg cagacatagc taatgcaaag 420
gcacggagat acagactaca caaactagaa gaaacgtaga gtaagctgct ttgtctctt 479

```

<210> 10

<211> 2599

<212> ADN

<213> Homo sapiens

<400> 10

```

gaaatccagg ggaccaatgc attttatcac taaaactatt tttatataat ttaagaata 60
taccaaaagt tgtctgattt aaagtgtgaa tacatgattt ctactttca tgtaagggtta 120
tccacttttg ctgaagatat tttttattga atcaaagatt gagttacaat tatacttttc 180
ttacctaaag ggataaaatg tacttttgat gaatcaggga atttttttta agttggagtt 240
tagttctaaa ttgactttac gtattactgc agttaattcc ttttttggct agggatgggt 300
tgataaacca caattggctg atattgaaaa tgaaagaaac ttaaaagggt ggatggatca 360
tgattactgt cgataactgc agataaattt gattagagta ataattttgt catttaaaaa 420
cacagttgtt tatactgccc atcctaggat gctcaccttc caagattcaa cgtggctaaa 480
acatcttctg gtaaattgtg cgtccatatt cattttgtca gtageccagga gaaatgggga 540
tgggggaaat acgacttagt gaggcataga catccctggg ccatccttcc tgtctccagc 600
tgtttcttgg aacctgctct cctgcttggc ggtccctgac gcagagaccg ttgctcccc 660
cacagccgtt tgactgaagg ctgctctgga gacctagagt aaaacggctg atggaagttg 720
tgggaccacac ttccatttcc ttcagtcatt agaggtggaa gggagggttc tccaagtttg 780
gagattgagc agatgaggct tgggatgccc ctgctttgac ttcagccatg gatgaggagt 840
gggatggcag caagtggtg cctgtggcag tggagtgtg ccagaaacag tggccagttg 900
tatcgccat aagacagggt aaggtctgaa gagctgagcc tgtaattctg ctgtaataat 960
gatagtgtc aagaagtgcc ttgagttggt gtacagtgcc atggccatca agaattccag 1020
atctcaggtt ttattacaaa atgtaagtgg tcacttggcg atttttagt acatgcatga 1080
gttacctttt ttctctatgt ctgagaactg tcagattaaa acaagatggc aaagagatcg 1140
ttagagtgc caacaaaatc actatcccat tagacacatc atcaaaagct tattttttatt 1200
cttgactgg aagaatcgta agtcaactgt ttcttgacca tggcagtgtt ctggctccaa 1260
atggtagtga ttccaaaata tggttctgtt aacactttgg cagaaaatgc cagctcagat 1320
atcttgagat actaaggatt atctttggac atgtactgca gcttcttgtc tctgttttgg 1380
attactggaa taccatggg cctctcaag agtgctggac ttctaggaca ttaagatgat 1440
tgtcagtaca ttaaactttt caatcccat atgcaatctt gtttgtaaat gtaaacttct 1500
aaaaatatgg ttaataacat tcaacctgtt tattacaact taaaaggaa ttcagtgaat 1560
ttgtttttat tttttaacaa gattttgtgaa ctgaatatca tgaacctgt tttgataccc 1620
ctttttcacg ttgtgccaac ggaatagggt gtttgatatt tcttcatatg ttaaggagat 1680
gcttcaaaat gtcaattgct ttaaacttaa attaccttc aagagaccaa ggtacattta 1740
cctcattgtg tatataatgt ttaatatattg tcagagcatt ctccagggtt gcagttttat 1800
ttctataaag tatgggtatt atgttgctca gttactcaaa tgggtactgta ttgtttatat 1860
ttgtacccca aataacatcg tctgtacttt ctgttttctg tattgtattt gtgcaggatt 1920

```

```

ctttaggctt tatcagtgtg atctctgcct ttttaagatat gtacagaaaa tgtccatata 1980
aatttccatt gaagtcgaat gatactgaga agcctgtaaa gaggagaaaa aaacataagc 2040
tgtgtttccc cataagtttt tttaaattgt atattgtatt tgtagtaata ttccaaaaga 2100
atgtaaatag gaaatagaag agtgatgctt atgttaagtc ctaacactac agtagaagaa 2160
tggaagcagt gcaataaata tacatttttc ccaagtgccg gtggcatatt ttaaaataaa 2220
gtgtatacgt tggaatgagt catgccatat gtagttgctg tagatggcaa ctagaacctt 2280
tgagttacaa gagtcttttag aagttttcta accctgccta gtgcaagtta caatattata 2340
gcgtgttcgg ggagtgccct cctgtctgca ggtgtgtctc tgtgcctggg ggctttttctc 2400
cacatgctta ggggtgtggg tcttccattg gggcatgatg gacctgtcta caggtgatct 2460
ctgttgccct tgggtcagca catttgtag tctcctgggg gtgaaaactt ggcttacaag 2520
agaactggaa aaatgatgag atgtggtccc caaaccttg attgactctg gggaggggct 2580
ttgtgaatag gattgctct
2599

```

<210> 11

<211> 222

<212> ADN

<213> Homo sapiens

<400> 11

```

gtaggagaga agtatggcta cctttcactt ccaaagccaa tggtaagtag aagaattatg 60
tacattgcag aatgttttta ttcagaattc aaattatgct ctgttgctctc 120
ccctctacca caaataattg acatgacaga aaaaagttag gatattaatg ttgacttcag 180
agttgttatt ctgggtagca atctagtgtt ggcttacact ct
222

```

<210> 12

<211> 3422

<212> ADN

<213> Homo sapiens

<400> 12

```

gaagtcactg ataattataa ttagtttaaa tttattctta cactacttaa gaaaaatttc 60
atttagaaaa ggctaccatg agatcttttg actatatgga gtagactgca tacagatctc 120
tttatttcta ttatgaatat tttagcgatt ataatcagct attccactga gtattagatt 180
tggtcactaa gtaactcatg tttttcttta gtcaatcatt cattcattca ctaaaacatt 240
tggtgagagt tggctaaact ctgattctg tgccaggcac tgtacagggg gatggagatc 300
actcataaga aggcctcttg ctcttagcca gatcatatgt aatactaaat ttaaattttt 360
tgcattaaaa ggctaaaaaa tcatttgttt ttgttaatta tatactgggt gatttataga 420
tgtaattagc acatattcat aagagacaaa tatatacaat gagttactat acatatacaa 480
agtgatttta gctgtataaa gtgatattgc aaaaaactac agctaagtgt tcttgctggg 540
taattattag cataaatatc aggttcttag aggtctttaa gagaataact tagagctacc 600
aaaattagca gcagtgatgt aatatcacc tgaaatacca ggagcagagg tgccccagag 660
tgcttcacag ctatgatcag gaaacttgac ctgccccatt gctacagcaa cggcataagg 720
actaggaagg tccactagct ggggaggtct tgggaacctc cgagccagaa agaagcattc 780
cagcatgatt cctcagatgc aatgaggaag gtttacagaa ttttcacagg gcaacagaga 840
tcaagtggta tgaaaaggta ggagttttaa gctcaagtct caaagataac ctacctttac 900
attctcagtc agaatttgac ccagaaagag gaagtttctt caggaacaga ccgtacgttg 960

```

```

atgaagtaga agtcatttca gtaaaaacgg agtacagaac ttgggtcataa tatcttgcac 1020
tttatagatt tattaaagat tagtttcaag ttcacattcg ctattcagtt gtaaaccgaa 1080
tggatgggag gggagaaaat acaagctctc cacacaggta tgctcctctc ttttctgaga 1140
gagaaggcat gggattttca gcataaattc catgttatgt gagtgctgtt tgagttctga 1200
agttcctatc aatatctgtt cctgcaagtg atctctgtaa gaccacctta catgctggctc 1260
ttagttattg ttaaaattgc aagggtttctt cacacctctt ttgataagaa gtgttttagct 1320
ggcagagctt tcccttgact tctgagtcta gtgtgggttg gcccatgaca gtgggaagaa 1380
atccaacatg ttacatggag accttgtatg taaacaaact ctgtagcctt tgaaagtgga 1440
actgcttttt acagttaaaag ggctgctaaa tggcttgacg atgagatctt ctggctcacc 1500
ttgatcttca catgaacca ttgtgacctt tctggatttc ctaggacctg tagtttccat 1560
ttgggttata ttagtgcttc aggaatgtgt cactactggt caagcaatac tcagaaattc 1620
aggctgtgag gggttagatt agaggaaagt aatttaggtg agctatccag agctttcaac 1680
ctcagggttt ataggaaaaa aaaacaaaat acaaaaaccc ttaggacttc tcatactcta 1740
gccctgacct tgcagtctcc tctgaggggc ctctaagac ccttaaaata atattacaat 1800
aggactttga gtgtgcattt cacatgtgga aatacctctg gctagcatta cttccaaatg 1860
ctagttccag ctaaaaacaa gtacttttta ttagtttctt gtctttcaac attctgccat 1920
caaaatatat tcattttatt aacttaccga ttattgggaa cctgcaaaaa aaaaaaaaaa 1980
aacacacatg ctgtctgcag ccactgtttg gttatttttg atagtttaag agaattataa 2040
atcaagttgt tctgtctact agagtcagaa attctgtgct agaatgggac ttctaaaatt 2100
agggtggctc gtactttctt agagaaagac ccaaattcat aggtttaagt ggcttgctcg 2160
acatacacat acagccagtt agtaacagaa ttcagaatag aatagagttc ttttttttct 2220
gtcagtcctg tgttctactt ctctttcctt ttcttcatgg ttgagcagct ttggcaaaa 2280
tgtgttccat atagcagtag ttccaagaga tgttaataga tgaagaaaaa aaagtgtgtg 2340
tgtgttgggg gtgggggatt gcatgggtcaa ttaaaatgaa ggaatgctga gtttgtttcc 2400
ttactgtagg acttctaaga gttttgaaca tgctattaca cattgtgaat ttataaaagg 2460
ggacaaacac tgtttgacct tagactgcac tccccttctt tattectcag agcatctcaa 2520
aggactacag ggtcctttga gaaacatggc ccttgatcaa gtttacaact ctttgagaag 2580
acatccctaa tggaaatact gaatggcaga tctattactg tggtagcatg tagtgtgtaa 2640
taaattctga agtccagaaa gcaagtatat ttgctgttta gagctatact taaggatatag 2700
agattctttt ccacctgtgg tgtgtaaaat cagtattgtg tttaattttg cagggtttttc 2760
aaggatcaac taatgtgcaa atacagtggg gatataattg tattacattt taataatgct 2820
tttgatggc actttatagt ttccagacac cttttgtaac agccctgtga gacagacaat 2880
attatcctta ttgtacagat gggggggctc tggaggctct gaaaagtttt tggatgatt 2940
cccctaagtt tataccacaa aatgatggag ccagaattta gttgcagAAC ttctgactcc 3000
aagtcagag ttctttacta gacatctggt ttcttactag ctttttgcta gctttttatt 3060
ttacagtga aagtatggaa gcaaatcctg ataattttg cttattatca aatatcattt 3120
attttttatt gcagagaccc agcagatgta actctcattg tagccatccc aaagaaaacc 3180
cttctttgcc ctttgacgtg ccacttattt aggttaactga agtgtatttc cagtacaatt 3240
tgattttgaa ctaagatttg atatagagtt gtaattgctt acataactct ttgaaaatgt 3300
atcatccctg ggttctcagg tgaaggaggc aattcaaaag taaaaagaa tattcgaaata 3360
caataaatac cctctgttct cacacaaatg ctgtttctta gctaaacttg tttgacaacc 3420
ta
3422

```

<210> 13

<211> 5791

<212> ADN

<213> Homo sapiens

<400> 13

```

ctaataattat tggccagtct gtggaagaag ctatcagagg ggtgtttgat gcttccctca 60
aaatggctgg cttctatgga ttgtatacct ggctgactca tactatgttt ggcatcaata 120
ttgtcttcat accatcagca ttagcagcaa tccttggagc agtgccattc ctggggacat 180
actgggcagc agtacctgca gttcttgacc tgtggctgac acaaggggta ggatgcaagg 240
ccattttact gttgattttt catctcttgc caacatactt tgtagatact gcaatctact 300
ctgacatatc aggaggtggc cacccttacc tgacaggctt ggcagtggcc ggtggagcat 360
actacctagg cctggaagga gcaatcatcg gtcctattct tctctgcata cttgtgggtg 420
cttccaatat ctatagtgc atgctagtga gtcccacgaa ttcagttccc acgccaacc 480
agaccccatg gctgctcag cctcagcgga ctttcctgta ctttctgaa gatctgaaat 540
cttcagttag ttgatgtggt ttctctgca gtgatttttc taggaagttc aaatttgaca 600
gagagttcag ctgagctgtg gccntctgcc ctccagctg tgcctagcaa gcaaaaccca 660
ggaaagaagc agaagcctcc tggccttaca tacagaatgc ctggacaaga gagaacttgc 720
tgccggctgc tttgtatttt aaaacacagc ttgagagttc agagtgggtg gtttgcctac 780
ttaactgttg ttaagatggc ttgaaaagtt tcattttata cactggtacc ctggcttgaa 840
atttttccac tttggttacc tatgttacta tattatataat ttataaagtt attttaagaa 900
ctctaaacta cctgctgtta aaagaataga tgggtgaatt ttttcctggt ttaagaaatg 960
tattgttaaa cttttctaag acagtcactt ttcaaggaag agggctttca cttttgagtg 1020
tgtagttgag tgagcaggaa aaatgaatct tctacccttc tcccacaatg tattatacgc 1080
tctttaagaa ataataaatc ataagtataa ggggtggggtg gcttatttga tgttcagttt 1140
tatattataa ccctgggaga taaaagact gaagctcttc ctccctcttc ttttctcttc 1200
acgcttcttt acactattgc caaattataa aacttggctg accacgttga agtgaaatac 1260
ttattaagct gctatgaatg gtaacagtat gataaaattc atgctgttat taggttttct 1320
ccttccaggt gggtgaagtg aaaaatctca ggtgtagcaa tactctgatt tgttaaagt 1380
attcgtttca ttatttgaa ctctgccaat aagttagttt tcagcagaat tttgtattta 1440
tgtagtattt tcccactct tagcacagt ccttgccat gactctaat gaatgcttgt 1500
tgaattgaat tgcagcttta aaaatatctc agatatcaca ataaccagca tagtgcttta 1560
cattttacaa agtccttttg cacatactat ttcacaagta catcacatca gctttcatat 1620
agctttatat tattttacat cagttttaca ctgatggaga agctaagggt cagaaaagt 1680
accaactcaa agtcgtaaag ttagtagga tttggcctca gatctttgac tccgttgtgt 1740
gcagattctg ctgtattagt gttccttcta aatattgtat gcagttttcc cagtgttang 1800
agattggtgt tgctatccat ttgttgtaa tgaagacttg tcctatagta catgacagg 1860
ctgagagtgg actgtagttc ttagctgaac ctttctact gtttaaaaca atcatcaggc 1920
ctatcctctg ctcccagggg aatgtgggag agaaaaggg tcttggccag gctgtagggt 1980
gtggctcaca tctgtaatcc cagcactttg gaaggctgag gtgggtggat tgcttgagcc 2040
cagggatttg agatcagcct aggcacatg gcaaaacccc atctctacaa aaaaaaaaaa 2100
aaaaaaaaaa aaattagctg ggcacggtg tgtggctata gtctgagcta ctcaggaggc 2160
tgatgtatga ggatcacctg agtcaggag gtcgaggctg cagtgaacca agatggcgcc 2220
actgcactcc agcctgggtg agagagtga acactgtccc aaggaaaaaa aaaagtgttg 2280
gggggtagtt cttaaagcgt gaagatttcc tttttggaaa gtatttaaag gtagaaatat 2340
taataatttt tcctaaaaca tctccctaag aagtcactga taattataat tagtttaaat 2400
ttattcttac actacttaag aaaaatttca tttagaaaag gctaccatga gatcttttga 2460
ctatatggag tagactgcat acagatctct ttatttctat tatgaatatt ttagcgatta 2520
taatcagcta tccactgag tattagattt ggtcactaag taactcatgt ttttctttag 2580
tcaatcattc attcattcac taaaacattt gttgagagtt ggctaaactc tagattctgt 2640
gccaggcact gtacaggggt atggagatca ctcataagaa ggctcttgc tcttagccag 2700
atcatatgta atactaaatt taaatttttt gcattaaaag gctaaaaaat catttgtttt 2760

```

tgtaattat	atactggttg	atctatagat	gtaattagca	catattcata	agagacaaat	2820
atatacaatg	agttactata	catatacaaa	gtgatttttag	ctgtataaag	tgatattgca	2880
aaaaactaca	gctaattgttt	cttgctgggt	aattatttagc	ataaatatca	ggttcttaga	2940
ggctctttaag	agaatacttt	agagctacca	aaattagcag	cagtgatgta	atatcaccct	3000
gaaataccag	gagcagagggt	gccccagagt	gcttcacagc	tatgatcagg	aaacttgacc	3060
tgccccattg	ctacagcaac	ggcataagga	ctaggaagggt	ccactagctg	gggagggtctt	3120
ggaaccctcc	gagccagaaa	gaagcattcc	agcatgattc	ctcagatgca	atgaggaagg	3180
tttacagaat	tttcacagggt	caacagagat	caagtgggtat	gaaaaggtag	gagttttaag	3240
ctcaagtctc	aaagataacc	tacctttaca	ttctcagtc	gaatttgacc	cagaaagagg	3300
aagtttcttc	aggaacagac	cgtacgttga	tgaagtagaa	gtcatttcag	taaaaacgga	3360
gtacagaact	tggtcataat	atcttgcaat	ttatagattt	attaaagatt	agtttcaagt	3420
tcacattcgc	tattcagttg	taaaccgaat	ggatgggagg	ggagaaaata	caagctctcc	3480
acacagggtat	gctcctctct	tttctgagag	agaaggcatg	ggattttcag	cataaattcc	3540
atgttatgtg	agtgtgtttt	gagttctgaa	gttcctatca	atatctgttc	ctgcaagtga	3600
tctctgtaag	accaccttac	atgctgggtct	tagttattgt	taaaattgca	aggtttcttc	3660
acacctctct	tgataagaag	tgtttagctg	gcagagcttt	cccttgactt	ctgagtctag	3720
tgtgggttgg	cccatgacag	tggaagaaa	tccaacatgt	tacatggaga	ccttgtagtg	3780
aaacaaactc	tgtagccttt	gaaagtggaa	ctgcttttta	cagttaaagg	gctgctaaat	3840
ggcttgacga	tgagatcttc	tggtcacct	tgatcttcac	atgaacccat	tgtgacctat	3900
ctggatttcc	taggacctgt	agtttcatt	tgggttatat	tagtgcctca	ggaatgtgtc	3960
actactggtc	aagcaatact	cagaaattca	ggctgtgagg	ggtagatta	gaggaaagta	4020
atthaggtga	gctatccaga	gctttcaacc	tcagggttta	taggaaaaaa	aaacaaaata	4080
caaaaaccct	taggacttct	catactctag	ccctgacctt	gcagtctcct	ctgaggggcc	4140
tcctaagacc	cttaaaataa	tattacaata	ggactttgag	tgtgcatctc	acatgtggaa	4200
atacctctgg	ctagcattac	ttccaaatgc	tagttccagc	taaaaacaag	tactttttrt	4260
tagtttcttg	tctttcaaca	ttctgccatc	aaaatatatt	cattttatta	acttaccgat	4320
tattgggaac	ctgcaaaaaa	aaaaaaaaaa	acacacatgc	tgtctgcagc	cactgtttgg	4380
ttatttttga	tagtttaaga	gaattataaa	tcaagttgtt	ctgtctacta	gagtcagaaa	4440
ttctgtgcta	gaatgggact	tctaaaaatta	gggtgtcctg	tcacttttcta	gagaaagacc	4500
caaattcata	ggtttaagtg	gcttgtctga	catacacata	cagccagtta	gtaacagaat	4560
tcagaataga	atagagttct	ttttttcttg	tcagtcctgt	gttctacttc	tctttccttt	4620
tcttcattgg	tgagcagctt	ttggcaaaat	gtgttccata	tagcagtagt	tccaagagat	4680
gttaatagat	gtaagaaaaa	aagtgtgtgt	gtgttggggg	tgggggattg	catggccaat	4740
taaaatgaag	gaatgctgag	tttgtttcct	tactgtagga	cttctaagag	ttttgaacat	4800
gctattacac	attgtgaatt	tataaaagggt	gacaaacact	gtttgacctt	agactgcact	4860
cccttctctt	attcctcaga	gcactctcaa	ggactacagg	gtccttttgag	aaacatggcc	4920
cttgatcaag	tttacaactc	tttgagaaga	catccctaata	ggaaatactg	aatggcagat	4980
ctattactgt	ggtaccatgt	agtgtgtaat	aaatcttgaa	gtccagaaaag	caagtatat	5040
tgctgttttag	agctatactt	aaggatataga	gattcttttc	cacctgtgggt	gtgtaaaatc	5100
agtattgtgt	ttaatttttg	aggtttttca	aggatcaact	aatgtgcaaa	tacagtgggtg	5160
atatattgat	attacatttt	aataatgctt	ttgtatggca	ctttatagtt	ttcagacacc	5220
ttttgtaaca	gccctgtgag	acagacaata	ttatccttat	tgtacagatg	ggggggctct	5280
ggaggtctctg	aaaagttttt	ggtgatattc	ccctaagttt	ataccacaaa	atgatggagc	5340
cagaatttag	ttgcagaact	tctgactcca	agtccagagt	tctttactag	acatctgggt	5400
tcttactagc	tttttgctag	ctttttatct	tacagtgaag	agtatggaag	caaatcctga	5460
taattttggc	ttattatcaa	atatcattta	ttttttattg	cagagaccca	gcagatgtaa	5520
ctctcattgt	agccatccca	aagaaaaccc	ttctttgccc	tttgagtgct	cacttattta	5580
ggtaactgaa	gtgtatttcc	agtacaattt	gattttgaac	taagatttga	tatagagttg	5640

taattgctta	cataactctt	tgaaaatgta	tcacccctgg	gttctcaggt	gaaggaggca	5700
attcaaaagt	aaaaaagaat	attcgaatac	aataaatacc	ctctgtttctc	acacaaatgc	5760
tgtttcttag	ctaacttgct	ttgacaacct	a			5791

<210> 14

<211> 2615

<212> ADN

<213> Homo sapiens

<400> 14

acccaaactt	caaaactggt	caataattaa	caatgtatta	atthttgaaaa	gttttttaact	60
gtaagagtat	caaataataat	acagagctct	taaataactta	aagctttctg	tcacaatatt	120
aaaaataaaa	cttcttagca	ttccaagact	atgaaaagtg	ccccctcca	atthttttaag	180
caagtaatag	tatctgctgg	tgthttatatt	tgthtaaaaat	ttaaaataac	aacaacaaaa	240
aaaaaacttc	taacttgaat	gtgatgtgac	ttgagtactt	ctaaaccatc	gaaaatacct	300
gatacgggga	gtaggttcat	ggthtttgga	aaagaaacta	tgacacagag	ttttaatatt	360
ataactggac	ttctaaatca	taaagtaaga	aaattaacat	catttagtat	ctthttcttgt	420
ttgthttttta	thtttgctaag	aaggaaccag	atthttatctc	atcattcttc	atctaaatgg	480
actgccatat	gaagttctgg	thttctctgtt	ctthttctgg	tagctgggta	ctaccaaatg	540
cctaaggaca	attctcagag	catttccaca	tgaaggtgat	tagthttcat	ttataaaaa	600
taataacaca	ttacatgtag	atgaaaatat	aacaaaacca	tacgaaagct	tgtagagat	660
attcaactag	aaaactaaaa	cagtagccac	tagaaaaaag	acatagtggg	aaaatthttg	720
aacggatatg	ctthttgttg	aataattaca	gtaattacta	ccaatatatt	cagaccaaca	780
aagattagaa	agaacatgca	aaatgctgag	aaaatthttc	ctacaattaa	aacaaatgtg	840
thtttatgag	agagaaaaaa	ataagtcaag	ctcctaaaa	gaaaattaca	aaacatagga	900
ttaacaattt	attaaaaatt	acacccttga	thttthttgag	gccccaaaaa	gtagthtttt	960
taaaacaagg	gcatgaaaaa	gacttcagat	taagactcat	actctgttct	aaatgaaaa	1020
tagtaaaact	acttctagta	gaaaatacct	gctatggaaa	aaaaatthtca	gatathtttc	1080
atthgatgta	ggtattaata	aaaaatatga	tgthttatctc	taaaatataa	actagcaatt	1140
tathtttcatt	tataatataa	agtgtthttg	gaaaagttht	ctcttctcag	tgthtcaaat	1200
aaaataggaa	tgthttctca	tacctagcaa	thtttaatgat	aatgtcaagt	aaaatthttaa	1260
thttgctgaa	ctaaagctca	aatagattca	aathtttaaaa	atatgaaaaa	tctacathtt	1320
taaaagctta	aaathtttgt	tactthtgcaa	tgacagatct	gctthtttcta	acathtttaaa	1380
aathttgtct	ctthtttatta	atataaaaact	attcaataaa	tcacaacaga	atctcttata	1440
ccggttaaagc	ataatcacat	gatacactcg	taacactcta	caaatacaat	ccagaaagg	1500
gaagtctccc	tcaaacaaag	aathttcttht	thttctgtcac	ctccattcca	tatatatagt	1560
cttcatthtca	catcacatac	atagaaaaat	tagcacataa	taacaggtag	ataagaaaaa	1620
aaathttttgg	ccttgtagag	ctcatcacgt	aaathttatc	atthttctta	gtatgtthtt	1680
thtttaacatt	atctgaggca	acatctaccg	gctthtgaggc	caatgaccca	acgccattgg	1740
aaggtgactt	tgcttcactt	thtttctcac	atgatgaaaa	cggctgtctgc	cagccaaata	1800
gcatccaact	caatgtatcc	tcaaaaggag	caaatgtcat	agtagaagtc	tgtgtagtca	1860
tcatgggttat	ctthtggtggc	acaataaaga	tggtataaa	gaacataaga	ttatgcttht	1920
atattgccag	atgctgggtg	tgctggaatg	tgaagataac	tgaccccgag	agggttacct	1980
tggcagggtta	tgacatcttc	agccattcat	cactthttctg	thttccaatct	ctthtttacc	2040
acctctcaaa	attcgttata	atgggtaccac	ctgcacctcg	ggcccatcca	atagctatca	2100
tgactatcca	gccathtttg	taatagctat	tagcatgtgt	gactccacct	actathtttc	2160
aagthtttgt	cacttctctc	attcccgaag	ccagtagttg	aacaggtaga	tatgaatagc	2220


```

cctgggaaac taggtcatgc gggcaaaaaa atgtaatat ccagattgaa gatgccagta 2280
atatgttagt gtggttttgc agaaacttca atggaggctc tgcaagcagt agacaggata 2340
aaattcctcc accaaaacag tggagcatag cagtaaacca gcttgaaata ggattcctcc 2400
atgccaatgc agctgtctcc ggctgacgtt tcaccgccat cactgacact agatagtgcg 2460
cgatgtcaaa aaaggggaaac atggacgtgc gggagaaggc cagagccaac tcgtcccatg 2520
gagaatccat aacgaccacc gaccgcggca accgaagcag ccgcggctac tggttccccg 2580
cgctcgcgcg gtgaggagta ggagagggca cagcc
2615

```

<210> 15

<211> 2551

<212> ADN

<213> Homo sapiens

<400> 15

```

ggctgtgccc tctctactc ctcaccgcgc gagcgcgggg aaccagtagc cgcggctgct 60
tcggttgccg cggtcggtgg tcgttatgga ttctccatgg gacgagttgg ctctggcctt 120
ctcccgcaag tccatgtttc ccttttttga catcgccac tatctagtgt cagtgatggc 180
ggtgaaacgt cagccgggag cagctgcatt ggcattggaag aatcctatct caagctgggt 240
tactgtctatg ctccactgtt ttggtggagg aattttatcc tgtctactgc ttgcagagcc 300
tccattgaag tttcttgcaa accacactaa catattactg gcatcttcaa tctggtatat 360
tacatTTTTT tgcccgcag acctagtttc ccagggtat tcatatctac ctgttcaact 420
actggcttcg ggaatgaagg aagtgaccag aacttggaag atagtaggtg gagtccacac 480
tgctaatagc tattacaaaa atggctggat agtcatgata gctattggat gggcccagg 540
tgcaggtggg accattataa cgaattttga gaggttggtt aaaggagatt ggaaaccaga 600
aggtgatgaa tggctgaaga tgcataccc tgccaaggta accctgctgg ggtcagttat 660
cttcacattc cagcacccc agcatctggc aatatcaaag cataatctta tgttccctta 720
taccatcttt atttgggcca caaagataac catgatgact acacagactt ctactatgac 780
atgtgtcctt tttagggata cattgagttg gatgctatct ggctggcagc agccgttttc 840
atcatgtgag aagaaaagtg aagcaaagtc accttccaat ggcgttgggg cattggcctc 900
aaagccggta gatgttgctt cagataatgt taaaaagaaa catactaaga agaatagaata 960
aatttacgtg atgagctcta caaggccaaa aatttttttt cttatctacc tgttatattg 1020
tgctaatttt tctatgtatg tgatgtgaaa tgaagactat atatatggaa tggaggtgac 1080
agaaagaaaag aaattctttg tttgaggagg acttcccctt tctggattgt atttgtagag 1140
tgttacgagt gtatcatgtg attatgcttt accggtataa gagattctgt tgtgattatt 1200
tgaatagttt tatattaata aaagaagaca aaatttttta aatgtagtaa aaagcagatc 1260
tgtcattgca aagtaacaaa aattttaagc ttttaaaaat gtagattttt catattttta 1320
aaatttgaat ctatttgagc tttagttcag cagaattaaa tttttacttg acattatcat 1380
taaaattgct aggtatggag aacaattcct attttatatt gaacactgag aagagtaaac 1440
ttttcctaaa acactttata ttataaatga aaataaattg ctagtattata ttttagatat 1500
aaacatcata ttttttatta atacctacat caaatggaaa atatctgaaa tttttttcc 1560
atagcaggta ttttctacta gaagtagttt tactactttt catttagaac agagtatgag 1620
tcttaatctg aagtcttttt catgcccttg ttttaaaaaa actacttttt ttggcctcaa 1680
aaaaatcaag ggtgtaattt ttaataaatt gttaatccta tgttttgtaa ttttcatttt 1740
aggagcttga cttatttttt tctctctcat aaaaacacat ttgttttaat tgtaggagaa 1800
attttctcag ctttttgcag gttctttcta atctttgttg gtctgaatat attggtagta 1860
attactgtaa ttattcaaca aaaagcatat ccgttcaaaa atttttccac tatgtctttt 1920
ttctagtggc tactgtttta gttttctagt tgaatatctc tgacaagctt tcgtatgggt 1980

```

```

ttgttatatt ttcactaca tgtaatgtgt tattaatttt attaatgaa aactaatcac 2040
cttcatgtgg aaatgctctg agaattgtcc ttaggcattt ggtagtaacc agctaaccac 2100
gaagaaacag agaaaccaga acttcatatg gcagtcatt tagatgaaga atgatgatat 2160
aaaatctggg tccttcttag caaaataaaa aacaaacaag aaaagatact aaatgatgtt 2220
aattttctta ctttatgatt tagaagtcca gttataatat taaaactctg tgacatagtt 2280
tcttttacca aaaccatgaa cctactcccc gtatcaggta ttttcgatgg tttagaagta 2340
ctcaagtcac atcacattca agttagaagt ttttttttgg ttgttgttat tttaaatttt 2400
taacaaatat aaacaccagc agatactatt acttgcttaa aaaattggga gggggcactt 2460
ttcatagtct tggaatgcta agaagtttta tttttaatat tgtgacagaa agctttaagt 2520
atttaagagc tctgtattat atttgatact c 2551

```

<210> 16

<211> 775

<212> ADN

<213> Homo sapiens

<400> 16

```

tctatccgct cgcgtcagcg ccttgccacc ctcatctcca atatgcctgg tccgaccccc 60
agtggcacta acgtgggac ctcagggcgc tctcccagca aagcagtggc cgcccgggcg 120
gcgggatcca ctgtccggca gaggtaagga accctgcagt tcgttcgctt ccagactcgg 180
agataggacc cagaacctcg ctgattctgg ggtggagacc ctagcatgtg aagattgaca 240
aaggcaaaat gagcttctag tgacgtggcc gtgggagtag ttaaaggcct tttgggagga 300
aggcgacatt tttttctcg ttgctcagtt tagggcacta ctcttaaaaa aggaaagtta 360
acaaactgga atagagtcag agataacttt gagaaaaccg atgtcattaa actggtgtct 420
ctggacctga ggtttgact cacatttcca tctggcggcc ccataagcaa tctgtcctac 480
agataactcg tcctacacaa aacttagtct cttttcagct cagctctctc actctcaatt 540
atatctcctt acttccatat ggcaactgtt tacactcatt tactcagagc cagaaacgct 600
agcgtcatct tggatttttc ttatgctctt tctctctcta gtcatatgcc agacttttaa 660
ctctgcttga aagctttctc ataagctctt tccttttccc tttctactgc tttgcatttg 720
ctacttaacc cttttcttca ggctgtttgc tttccagtc atcgttcgct ctgct 775

```

<210> 17

<211> 491

<212> ADN

<213> Homo sapiens

<400> 17

```

ccctcctttc ccaccagaca ctcccttaca ccatgcacat ttatttaatn atgtacttaa 60
ttcaaactta acaccagggc aacacttata ctaaactatg tgctttcctt aagaagttcc 120
caggggcta atgtggaagaga aaaccaggcc tggagatcc agtttcaaaa atnccagaga 180
ttacctcaa ggccgtttgt caacaacccc agctattact gagatggagg ccaccatgct 240
cnnnggtggg accacggact caaggatagc cactggaaac aaggacacag aggacctgt 300
gccaggcaa cactcccgcg tgccctccat tccaagggtt cttttttang cccctctctt 360
ccagccaaag ttgnaaatgg ggtttnttta nggcactagc cttgggntat ttttaccn 420
tgnnngctnt ggnagggaan gtcantttcc tttcactggn tctcaccttt gttacctggc 480
tttgcaagtg g 491

```

<210> 18
 <211> 293
 <212> ADN
 <213> Homo sapiens

<400> 18
 caccatgcac atttatttaa ttatgtactt attcaactta caccaggcac acttatctaa 60
 ctatgtgctt tcttagaagt tccagggcta atgtggagag aaaccaggcc tggagatcca 120
 gtttcaaaat tccagagatt acctcaaggc cgtttgtcaa caaccagct attactgaga 180
 tgaggccacc atgctccagg tggaccacga ctcaagatag ccactgaaac aagacacaga 240
 gaccttgtgc ccagcaacac tcccgcacgc ctccattcc aaggttccct ttt 293

<210> 19
 <211> 2442
 <212> ADN
 <213> Homo sapiens

<400> 19
 tatgtgggat tttagggcaa ggattttcaa cagtcaacag tactggaagc tgaggcttca 60
 gaacccagggt atggaaggca gggagtttga actatgctcc ttgcacagaa tgggaagcca 120
 ggaaaatgca aatgtcaaac tatectacag ggaagtctct ataaacctcc acagaactcc 180
 cacaaaagac acacttaggg aaaaggatat cagcccaaaa tatacaagtc tgttcaactt 240
 gctccctaaa ataagggaga atgccgacac cgcaggaaaa aggagctaat ggaactaaca 300
 agaactttaa agaaacacca ataaaccaag aacacgcgaa aatgacaaaa taccataata 360
 aaatccaaaa tagaaacaaa tccaaaagaa aacaaaagtt gatgaaataa aatctcagta 420
 ggaggattaa gtaaaagaat tagcaaatg aaagattttt aagagacgaa gacagaataa 480
 aaagctccct atgtttaaaa ggagttttta gaagaaaata gaaacaatgg ggggaagatgg 540
 ttaaaaatta ataataactt tccagacttg aagaaaaaaa actgaatctt cacattacag 600
 aagcacacta tgtctagggc aggataaata acaaatctac actaggacac atcctgatct 660
 gcaacacact gaagataaag agagaacagt aaaaatatag attactacct gtaagattac 720
 tacctacaaa ctaagagggc tcagactaac agcagatttt ttcattagtt aacactgagg 780
 caaggagaca atagaacagt atcttcaagg cacagcagac agtactaaac aactcaactc 840
 ggtcaaacct aactttcttg ccagcacttg tgaaatcaaa ggctccccag taatttgatt 900
 agcactttca tcatttttga atcttataaa cactttctcc agtggggaag aaatagccca 960
 aatgggaatt ttaggtacct aagacttcgg ctgccactg tgccataat cctagtaatg 1020
 ttaacctttc cagtacccag tatagaatac ccctacaact gggcatttta aacaggcaaa 1080
 caagtattac acacacttaa ttttgttcat tccagccaca tgaagtcaaa ggatgtttct 1140
 ctataagaaa ctacaataga tctttgaaac tttgcatgaa gttgttttag gggccatact 1200
 tcaagtgaca ggaaataaca aatagcaata cacacacatt tcattaagtg gggatcttat 1260
 aaaatctcta gaacctgag aagttccttg ctacttataa agccttttct tgacctgtca 1320
 gacctaatgt ttgatgatgt ataagaaaac tatcgtcgaa tttatgttat gcataaattt 1380
 tatagtatat ttaaagtga atacacaatt tttttctatg cacatactct tactggaaaa 1440
 aatttgaatg ctgacaataa aatggctaag ttatggtaca tgcagtgaac agaatactat 1500
 agagcagact gagcaaacta tggccacagg ccaaatccag cccatggtaa gtatttgtaa 1560
 atagttttac tggaacacaa ttatactcat ttatttacat attatctatt gatgctttca 1620

```

cactaccgtg gtagtgctga gtagccatga cagagaccta atggtctgca aaacctaata 1680
tatttactaa ctaacccttt atagaaagaa ctttgctgaa ccctgggtatg gagaaataaa 1740
aacgaatgaa atacagctga atgagtcac ttcaatcaat ctcaaaagca atgctgagag 1800
aatcaaatca caaaaaatac atgcagtata attcccatta caaaaaatcc acaaatatgc 1860
aaaacaaacc tgatactgtt cacaatgca cataaaacca tttttttttt taaggtaatg 1920
aagttgatca acaccaagtg atctggggca gagggtacg tagaagacat ttaactgtgc 1980
agggtagatt aacagtatat gtgctgatga tattctatat ccgaacctgg ggttggttaa 2040
atggaggttc agtttataat taccttttaa actacataaa tgttttttta tacattcagt 2100
atatgtattt tcatgtcatt taaaagttca aaaaagacca aaaaggcaca tatcaaaaca 2160
ttaacagcag gtgacagaat atttgccaac ctgttagcag gttttaaaac cttctacaat 2220
gaacttttaa tgggcacaca aataagaaca caagcataaa atacttttta aaattaaagt 2280
aaaatcttcc caactacgct aaaagctcta gcctgaatat tttatatacg agcaaaagta 2340
ctaattaaaa gcactagcat cttagaagaa tttccttaaa tccaagattt taaaattccc 2400
actctggcta gctctaatag agctcaaaac aaaatacaaa gg 2442

```

<210> 20

<211> 4608

<212> ADN

<213> Homo sapiens

<400> 20

```

ctcaaaacac actttccccc atgccccctg cataaattca ttcttatgct aatactgcaa 60
agcctccggt accatgcttc ttaccgtgat attcttgatg atctgctgca ccaagatcgc 120
attggtcaac atatgagtcg tgaggggtgc atgctgaagc agcagtcgaa gcagctggcc 180
cacaacaggc agctcctcag gcccaactct gtccaccgca aggctctgca gcatcaacct 240
gaggactcga ggcaagagag ccaggatgga gacacagacc cccacagct tgtccctaca 300
ataacagaga cagaaatgag tcagtgaagc aatgccaca ggaaaagccc ccttctata 360
ttcccccaa gtgaaatgtc cttttaagga acatgttttc caataacatt taaactgaga 420
gccaaatcaa gaacgcaatc tcacttacaa tagccacaca cactaaatac ctaggaatac 480
atctaacca ggaggtgaaa gatctctaca agaaaaacta cagaacactg ctgagagaaa 540
tcacagatga cacagataaa tggaaaaaca ttccatgctc atgaattgga ataatacagta 600
ttgttaaaat ggccacactg ctcaaagcaa cctacagatt caactgtatt cctatcaaac 660
taccaacatc attcttcaca caattaaaaa aaaatattct aaaattcaca tggaaacaaa 720
aaagccagaa tggccaaagc aatcctaagc aaaaacaacc tgacttcaca gtacactaca 780
aggctacagt agccaaaaca gcatggtaat gatagagaac aacagacaga taaaccaatg 840
gaacagaaaa gagaaccagc aaataaagcc acatacctac aactatctca tcttcaacta 900
aggtgacaaa aacaagcaac agggcaagga ttccctattc aataaatggt gctgggacaa 960
ttggctatcc tttttagtaa gaataaaaaat ggatccctat ctttcatcat atacaaaaaa 1020
ttaactcggc caggcgcggt ggctcacgcc tgtaatccca gcactttggg aggctgaggc 1080
aggtggatca cgaagtcaag aaatcaagac catcctggcc aacatggaga aaccccgctc 1140
ctactacaga tacaaaatta gccaggcatg gtggcacgcg cctgcagtcg cagctactcg 1200
ggaggctgag gcaggagaat tgcttgaatc cgggaggtgg aggttgtagt gagccgagat 1260
cacaccactg cattccagcc tggcaaaaga atgagactcc gtctcaaaaa aaaaattaac 1320
tcaagataga ttaagaattc attactaagt cctcaaaagc aattgcaaaa acaaaaaatg 1380
acaagtgggc cctaattaaa cttaagaaca agagaaacta tccacatttc aataaacatg 1440
gtgtttttca ctaacattta ggctaagagc caaatcaaga acatcattct ctcggtatct 1500
ttatggagca aaatgtttat ggtagagaca ggtgaccttg cagctagcct aaccacattc 1560

```

tccaagtgtt	aattagatgg	tgcaatattc	acatagacct	agagtccact	gactatgaaa	1620
ggctctgacc	tttttgcagc	ttctttacca	ggagtgtata	atcttcctag	ttcctaaatc	1680
tttctaagag	aacctgacta	cctccaggtc	tagagggagc	agccccttca	ctatataacc	1740
aatttccctg	actgcacata	tggtaatcc	atctatgggtg	aataatgaat	ccatttctagg	1800
ccaaatagat	actcttatgc	agagcccagc	tgggacagtc	acgtgcaagg	agaaaagaga	1860
aagccagact	gcagacatag	ctaatagcaa	ggcacggaga	tacagactac	acaaactaga	1920
agaaacgtag	agtaagctgc	tttgtctctt	gctaaatttc	taattcctag	cccttgggtt	1980
ttacaagtta	ctcctgtaat	aatcaaatat	tttcattttc	tacttttctt	aatttgagtg	2040
gatttcaa	aaccacagac	cttaaaaaaa	aaagacagcc	aagctcaagg	ccagatgggt	2100
aaaacctcta	ggtgaaagaa	actataatga	aattcatata	aagatatcca	acagtcataa	2160
aaactgtatg	tgggatttta	gggcaaggat	tttcaacagt	caacagtact	ggaagctgag	2220
gcttcagaac	ccaggtatgg	aaggcagggg	gtttgaacta	tgctccttgc	acagaatggg	2280
aagccaggaa	aatgcaaatg	tcaaaactatc	ctacagggaa	gtctctataa	acctccacag	2340
aactcccaca	aaagacacac	ttagggaaaa	ggtagcagc	caaaaatata	caagtctgtt	2400
caacttgctc	cctaaaataa	gggagaatgc	cgacaccgca	ggaaaaagga	gctaattgga	2460
ctaacaagaa	ctttaaagaa	acaccaataa	accaagaaca	cgcgaaaatg	acaaaatacc	2520
aaataaaaaat	ccaaaataga	aacaaatcca	aaagaaaaca	aaagttgatg	aaataaaatc	2580
tcagtaggag	gattaagtaa	agaattagc	aaattgaaag	atttttaaga	gacgaagaca	2640
gaataaaaaag	ctccctatgt	ttaaaaggag	tttttagaag	aaaatagaaa	caatggggga	2700
agatgggttaa	aaattaataa	taactttcca	gacttgaaga	aaaaaaactg	aatcttcaca	2760
ttacagaagc	acactatgtc	tagggcagga	taaataacaa	atctacacta	ggacacatcc	2820
tgatctgcaa	cacactgaag	ataaagagag	aacagtaaaa	atatagatta	ctacctgtaa	2880
gattactacc	tacaaactaa	gagggctcag	actaacagca	gattttttca	ttagttaaca	2940
ctgaggcaag	gagacaatag	aacagtatct	tcaaggcaca	gcagacagta	ctaaacaact	3000
caactcggtc	aaacctaact	ttctggccag	cacttgtgaa	atcaaaggct	ccccagtaat	3060
ttgattagca	ctttcatcat	tttgtaatct	tataaacacc	ttctccagtg	gggaagaaat	3120
agcccaaag	ggaatttttag	gtacctaa	cttcggctgc	caactgtgcc	taataccta	3180
gtaatgttaa	cctttccagt	acccagtata	gaataccctt	acaactgggc	attttaaaca	3240
ggcaaacaaag	tattacacac	acttaatttt	gttcattcca	gccacatgaa	gtcaaaggat	3300
gtttctctat	aagaaactac	aatagatctt	tgaaactttg	catgaagttg	ttttaggggc	3360
catacttcaa	gtgacaggaa	ataacaaata	gcaatacaca	cacatttcat	taagtgggga	3420
tcttataaaa	tctctagaac	cctgagaagt	tccttgctac	ttataaagcc	ttttcttgac	3480
ctgtcagacc	taatgtttga	tgatgtataa	gaaaactatc	gtcgaattta	tgttatgcat	3540
aaattttata	gtatatttaa	agtgaataac	acaatttttt	tctatgcaca	tactcttact	3600
ggaaaaaatt	tgaatgctga	caataaaatg	gctaagttat	ggtacatgca	tgtaacagaa	3660
tactatagag	cagactgagc	aaactatggc	cacaggccaa	atccagccca	tggttaagtat	3720
ttgtaaatag	ttttactgga	acacaattat	actcatttat	ttacatatta	tctattgatg	3780
ctttcacact	accgtggtag	tgctgagtag	ccatgacaga	gacctaatgg	tctgcaaaac	3840
ctaaaatatt	tactaactaa	ccctttatag	aaagaacttt	gctgaacctt	ggtatggaga	3900
aataaaaacg	aatgaaatac	agctgaatga	gtcaacttca	atcaatctca	aaagcaatgc	3960
tgagagaatc	aaatcacaaa	aaatacatgc	agtataattc	ccattacaaa	aaatccacaa	4020
atatgcaaaa	caaacctgat	actgttcaca	aatgcacata	aaaccatttt	tttttttaag	4080
gtaatgaagt	tgatcaacac	caagtgatct	ggggcagagg	ggtacgtaga	agacatttaa	4140
ctgtgcaggg	tacattaaca	gtatagtgtc	tgatgatatt	ctatatccga	acctgggggt	4200
ggttaaatgg	aggttcagtt	tataattacc	ttttaaacta	cataaatgtt	tttttataca	4260
ttcagtatat	gtattttcat	gtcattttaa	agttcaaaaa	agacaaaaaa	ggcacatatc	4320
aaaacattaa	cagcaggtga	cagaatattt	gccaacctgt	tagcaggttt	taaaaccttc	4380
tacaatgaac	ttttaatggg	cacacaaata	agaacacaag	cataaaatac	tttttaaaat	4440

```

taaagtaaaa tcttcccaac tacgctaaaa gctctagcct gaatatttta tatacgagca 4500
aaagtactaa ttaaaagcac tagcatctta gaagaatttc cttaaatcca agatttttaa 4560
attcccactc tggctagctc taatgaagct caaaacaaaa tacaaaagg 4608

```

<210> 21

<211> 1627

<212> ADN

<213> Homo sapiens

<400> 21

```

gaatgctgtt acctgaaccc cttacttcga aggatcataa gattcacagg ggtgtttgca 60
tttggaacttt ttgtactga catttttgta aacgccggac aagtgggtcac tgggcactta 120
acgccatact tcctgactgt gtgcaagcca aactacacca gtgcagactg ccaagcgcac 180
caccagttta taaacaatgg gaacatttgt actggggacc tgggaagtga agaaaaggct 240
cggagatcct ttcctccaa acacgctgct ctgagcattt actccgcctt atatgccacg 300
atgtatatta caagcacaat caagacgaag agcagtcgac tggccaagcc ggtgctgtgc 360
ctcggaaactc tctgcacagc cttcctgaca ggcctcaacc gggctctctga gtatcggaac 420
cactgctcgg acgtgattgc tgggttcatc ctgggcactg cagtggccct gtttctggga 480
atgtgtgtgg ttcataactt taaaggaacg caaggatctc cttccaaacc caagcctgag 540
gatccccgtg gagtaccctt aatggctttc ccaaggatag aaagccctct ggaaacctta 600
agtgcacaga atcactctgc gtccatgacc gaagttacct gagacgactg atgtgtcaca 660
agctgttttt taaaatcatc ttccaattct atacttcaaa acacacagtt gctcaatgtc 720
aaactgtgat gacaaatatt acgtttatct agttagaagc taatgttttg tacatttttt 780
gtatgaggaa gtgatgtagc ttgcctgat ttttttttt tttttggtca gctttaatat 840
atttatgcca gaatttttaa accaacaaaa ttttcttggt caagcgtgca ttgaagaacc 900
acattttattc aatggttgac gttgttttgt gatatttgta cacaaatttt cttttctcag 960
ttttataaac acagaatata acaattcact ttaacttttt attaccacag ttgctgcctc 1020
ctccagaatt tttgaatttt aataaaaggc aaacttttga gctgcaggaa ggacaatggt 1080
ggtaataaat aaatctcaaa gtcaattgta gaaaaaaaat tgtcttcaaa aagaatgttg 1140
cactctgata tcttaacaaa ttgttacgtt caaagtttaa agtgatata taacaaagtc 1200
acctagttat acaaacaatt gtcagagaat tctggatttg gagggattg gggttatatg 1260
attctttctt agataatggc ctctactaaa taactcaaga tctttctgga atgtctctctg 1320
gcaggcaggt gccactgtca gcttttctcc aaaaagcagc caacatcagc ctcccctgtc 1380
aactcaacag ttttgtatct catattatat ggactttata tgaaaatgaa tattttacag 1440
tttgcacagt attattttac agaaaaggaa tcagagaatc tacaacatag ggccccagaa 1500
caacagtttc actttgtggc ttttaattat tctagaattt taactgcac tcatttttct 1560
agcatggtga gaactaatat gtaactcctt tgattgaagg agctcttttg tccgtacct 1620
tcagaat 1627

```

<210> 22

<211> 2333

<212> ADN

<213> Homo sapiens

<400> 22

```

attcaagaga gcaaggacaa ataatgggtca aggaaggaag tcagggttg gtggctctgt 60

```

```

actaagtgtc agcttcaaag ggagagacca gagcataaac agacaattct ctaggttgca 120
aattgggagg attaggaaca agtaagagag ttccatgcag gatggacaag aaaaaaaagt 180
tttataccaa aaattgtatt tcatggcatt agggcagctc actggggtaa ggcactatgc 240
agtactaggc aattagaatt ctattcacta gctatgtgca caagcggctc ctgacctggg 300
cattttggct gtaaaactat tcatgcctgg acagtttttg acggtgcagt cttgctatat 360
gggtgtgagaa atggctgtag gaaacaacac tcaacgaagt tattccatca tcccgtgttt 420
tatatttgtt gagcttgtca tcatggctgg gacagtgcgt cttgcctact acttcgaatg 480
cactgacact tttcagggtc atatccaagg attccttctgt caggacggag acttaatgaa 540
gccttaccga gggacagagg aagaaagctt catcaccctc ctggtgctct atttgtgtgt 600
ggctgccacc ccaactgcta ttatttttat tggtagata tccatgtatt tcataaaatc 660
aacaagagaa tccctgattg ctcaggagaa aacaattctg accggagaat gctgttacct 720
gaacccttta cttcgaaggc tcataagatt cacagggttg tttgcatttg gactttttgc 780
tactgacatt tttgtaaacg ccggacaagt ggtcactggg cacttaacgc catacttcct 840
gactgtgtgc aagccaaact acaccagtgc agactgccaa gcgcaccacc agtttataaa 900
caatgggaac atttgtactg gggacctgga agtgatagaa aaggctcgga gatcctttcc 960
ctccaaacac gctgctctga gcatttactc cgcttatat gccacgatgt atattacaag 1020
cacaatcaag acgaagagca gtcgactggc caagccggtg ctgtgcctcg gaactctctg 1080
cacagccttc ctgacaggcc tcaaccgggt ctctgagtat cggaaccact gctcggacgt 1140
gattgctggt ttcacctctg gcactgcagt ggccctgttt ctgggaatgt gtgtggttca 1200
taactttaaa ggaacgcaag gatctccttc caaacccaag cctgaggatc cccgtggagt 1260
accctaagt gctttcccaa ggatagaaaag cctctggaa accttaagtg cacagaatca 1320
ctctgcgtcc atgaccgaag ttacctgaga cgactgatgt gtcacaagct gttttttaaa 1380
atcatcttcc aattctatag ttcaaaacac acagttgtct aatgtcaaac tgtgatgaca 1440
aatattacgt ttatctagtt agaagctaat gttttgtaca ttttttgtat gaggaagtga 1500
tgtagcttgc cctgattttt tttttttttt tggcagctt taatatattt atgccagaat 1560
tttaaaacca acaaaatttt ctgtttcaag cgtgcattga agaaccacat ttattcaatg 1620
gttgacgttg ttttgtgata tttgtacaca aattttcttt tctcagtttt ataaacacag 1680
aatataacaa ttcactttta actttttatta ccacagttgc tgcctcctcc agaatttttg 1740
aatttttaata aaaggcaaac ttttgagctg cagggaaggac aatgttggtt aataataaat 1800
ctcaaagtca attgtagaaa aaaaattgtc ttcaaaaaga atgttgact ctgatctctt 1860
aacaattgt tacgttcaaa gtttaaagt atataatca aaagtcacct agttatacaa 1920
acaattgtca gagaattctg gatttggagg gtattggggt tatatgatc tttcttagat 1980
aatggcctct actaaataac tcaagatctt tctggaatgt cttctggcag gcaggtgcca 2040
ctgtcagctt tttccaaaa agcagccaac atcagcctcc cctgtcaact caacagtttt 2100
gtatctcata ttatatggac tttatatgaa aatgaatatt ttacagtttg cacagtatta 2160
ttttacagaa aaggaatcag agaactctaca acatagggcc ccagaacaac agtttcactt 2220
tgtggctttt aattattcta gaattttaac tgcactcat ttttctagca tggtgagaac 2280
taatatgtaa ctcttttgat tgaaggagct cttttgtccg tacctatcag aat 2333

```

<210> 23

<211> 2764

<212> ADN

<213> Homo sapiens

<400> 23

```

cagtttttat cacaggtgca agccgtggca ttggcaaagc tattgcattg aaagcagcaa 60
aggatggagc aaatattgtt attgctgcaa agaccgcca gccacatcca aaacttctag 120

```

```

gcacaatcta tactgctgct gaagaaattg aagcagttgg aggaaaggcc ttgccatgta 180
ttgttgatgt gagagatgaa cagcagatca gtgctgcagt ggagaaaggcc atcaagaaat 240
ttggaggaaat tgatattctg gtaaaataatg ccagtgccat tagtttgacc aatacattgg 300
acacacctac caagagattg gatctgatga tgaacgtgaa caccagaggc acctaccttg 360
catctaaagc atgtattcct tatttgaaaa agagcaaagt tgctcatatc ctcaatatca 420
gtccaccact gaacctaaat ccagtttggg tcaaacagca ctgtgcttat accattgcta 480
agtatgggat gtctatgtat gtgcttgga tggcagaaga atttaaagggt gaaattgcag 540
tcaatgcatt atggcctaaa acagccatac acactgctgc tatggatatg ctgggaggac 600
ctggatcgga aagccagtggt agaaaagttg atatcattgc agatgcagca tattccattt 660
tccaaaaggcc aaaaagtttt actggcaact ttgtcattga tgaaaatata ttaaaagaag 720
aaggaataga aaattttgac gtttatgcaa ttaaccagg tcatcctttg caaccagatt 780
tcttcttaga tgaataccca gaagcagtta gcaagaaagt ggaatcaact ggtgctgttc 840
cagaattcaa agaagagaaa ctgcagctgc aaccaaacc acgttctgga gctgtggaag 900
aaacatttag aattgttaag gactctctca gtgatgatgt tgtaaaggcc actcaagcaa 960
tctatctggt tgaactctcc ggtgaagatg gtggcacgtg gtttcttgat ctgaaaagca 1020
agggtgggaa tgtcggatat ggagagcctt ctgatcaggc agatgtggtg atgagtatga 1080
ctactgatga ctttgtaaaa atgttttcag ggaaactaaa accaacaatg gcattcatgt 1140
cagggaaatt gaagattaaa ggtaacatgg ccctagcaat caaattggag aagctaata 1200
atcagatgaa tgcagactg tgaaggaaaa tataaaaaaa aagtcgactg ctatgctcaa 1260
aaagtaaaaa aagctcaaca gttaaaatct aatgtttgtt ttctttcctg ttatattata 1320
aggatatgca cgtttgttct ggaaaagata gaatttgtct ctaaaagact tgaaattgta 1380
attaaaatgg caagctaata aacataagc ttcattaatg gggattctaa gacagctgtg 1440
gtttttatat ttcaagggtt taaccctttg agccttacat ctcatcact gtctttctcc 1500
aagaaaagta ttttgggagg acagtcagat caagcagtaa aattagctct ttcaaactct 1560
cttgtcatgt aaaatgaagc tagtctgttt taaaattttt agttttggat tgtatactaa 1620
tgaaaatctt aatgatgttt ttgattttta tatacttatt ttaaagaaaa tcttatatag 1680
tacattttac aaaaattata aaaaatgaat tagtactggc gaggactaaa tgaaacaata 1740
atttttcatt ttgataacta gctttccagg tggacttagc cataggaaaa tattactaat 1800
gtaatttaac aaattgctgc atgtattcca tttaaaaata tgtttaaatt gtcctaaaac 1860
aaaaataattt tctccctagg agtatgcatt tggctacagt gttttgaaac agaaacctta 1920
gaatagggtca ttggtatggg ctgaactgtg tatcccccaa ttcatttgtt gaggtcctaa 1980
ctcccatctt ttttgaatgt gactgttcgg agatgaggcc tttaaagagg tgacttaagt 2040
tcaaaggagg ctgttagtct aatccaacat ggtgtccttt ggacataaga gataccagca 2100
atgtgtgcac agaacaaaga ccaggagagg acacagttag aaggcagtta tctgcaagca 2160
aagagagagg cttcagaaga aacaaaatca ccagcacctt gatctttgac ttctaactct 2220
cagaatagtg agaaataaat ttctgttgtt aagccgtcca ctgtgggagg ccgacgcagg 2280
aggattgctt gaggccagga gttcaaggcc agcctggaca acatagtaag accctatctc 2340
tcccccccta ataaattaat ttaaaaagcc ccccaatctg tggtatttta ttatggcagc 2400
cctagcaagc taatacagt gtttgagagg ctgggagggt tgaggggaag ataaaacttt 2460
aaaaagctct tatctttcat ttcaatcagt taaaaatact tgctcagtgt aacaattttg 2520
cttctcagct tccactctaa tattgttgtg ccattaagca atttagctaa tcttgacatt 2580
tcttagattc ataattgttag gagcatttaa tctgtatttt acaagttagg aagcagagga 2640
tcagagatgg gaaaggacta gcccaaggcc aacattaaca agccctctaa caaaaacttt 2700
acaatacatt tatgttgaat ggaactccaa gatctcacct ctccatccag gaatggagtc 2760
catg

```


<211> 3228

<212> ADN

<213> Homo sapiens

<400> 24

```
gctcgccgcc gccgctgtcg ccgccacctc ctctgatcta cgaaagtcac gttacccaac 60
accgggaggc tggcaggatg tacagttttt atcacagggt caagccgtgg cattggcaaa 120
gctattgcat tgaaagcagc aaaggatgga gcaaatattg ttattgctgc aaagaccgcc 180
cagccacatc caaaacttct aggcacaatc tatactgctg ctgaagaaat tgaagcagtt 240
ggaggaaagg ccttgccatg tattgttgat gtgagagatg aacagcagat cagtgtgca 300
gtggagaaag ccatacaaga atttgaggga attgatattc tggtaaataa tgccagtgcc 360
attagtttga ccaatacatt ggacacacct accaagagat tggatctgat gatgaacgtg 420
aacaccagag gcacctacct tgcactctaa gcatgtattc cttatttgaa aaagagcaaa 480
gttgctcata tcctcaatat cagtcacca ctgaacctaa atccagtttg gttcaaacag 540
cactgtgctt ataccattgc taagtatggg atgtctatgt atgtgcttgg aatggcgaga 600
gaatttaaag gtgaaattgc agtcaatgca ttatggccta aaacagccat acacactgct 660
gctatggata tgctgggagg acctggtatc gaaagccagt gtagaaaagt tgatatcatt 720
gcagatgcag catattccat ttccaaaag ccaaaaagtt ttactggcaa ctttgtcatt 780
gatgaaaata tcttaaaaga agaaggaata gaaaattttg acgtttatgc aattaaacca 840
ggtcatcctt tgcaaccaga ttcttcttta gatgaatacc cagaagcagt tagcaagaaa 900
gtggaatcaa ctggtgctgt tccagaattc aaagaagaga aactgcagct gcaacaaaaa 960
ccacgttctg gagctgtgga agaaacattt agaattgtta aggactctct cagtgatgat 1020
gttgtaaag ccactcaagc aatctatctg ttgaaactct ccggtgaaga tgggtggcacg 1080
tggtttcttg atctgaaaag caagggtggg aatgtcggat atggagagcc ttctgatcag 1140
gcagatgtgg tgatgagtat gactactgat gactttgtaa aaatgttttc agggaaacta 1200
aaaccaacaa tggcattcat gtcagggaaa ttgaagatta aaggtaacat ggccctagca 1260
atacaattgg agaagctaag gaatcagatg aatgccagac tgtgaaggaa aataaaaaa 1320
aaaagtgcag tgctatgtc aaaaagttaa aaaagctcaa cagttaaaat ctaatgtttg 1380
ttttctttcc tggtatatta taaggatatg cacgtttgtt ctggaaaaga tagaatttgt 1440
ctctaaaaga cttgaaattg taattaaaat ggcaagctaa tcaaacataa gcttcattaa 1500
gtgggattct aagacagtct gtgtttttat atttcaaggg tttaaccctt tgagccttac 1560
atctcattca ctgtctttct ccaagaaaag tattttgggc ggacagtcag atcaagcagt 1620
aaaattagct ctttcaaate ttcttgtcat gtaaaatgaa gctagtctgt tttaaaattt 1680
ttagttttgg attgtatact aatgaaaatc ttaatgatgt ttttgatttt tatatactta 1740
ttttaaagaa aatcttatat agtacatttt acaaaaatta taaaaaatga attagtactg 1800
gcgaggacta aatgaaacaa taatttttca ttttgataac tagctttcca ggtggactta 1860
gccataggaa aatattacta atgtaattta acaaattgct gcatgtattc catttaaaaa 1920
tatgtttaaa ttgtctaaa acaaaataat tttctcccta ggagtatgca tttggctaca 1980
gtgttttgaa acagaaacct tagaataggc cattgggtatg ggctgaactg tgtatcccc 2040
aattcatttg ttgaggtcct aactcccatt tcttttgaa gtgactgttc ggagatgagg 2100
cctttaaaga ggtgacttaa gttcaaagga ggctgttagt ctaatccaac atggtgtcct 2160
ttggacataa gagataccag caatgtgtgc acagaacaaa gaccaggaga ggacacagt 2220
agaaggcagt tatctgcaag caaagagaga ggcttcagaa gaaacaaaat caccagcacc 2280
ttgatctttg acttctaate tccagaatag tgagaaataa atttctgttg ttaagccgtc 2340
cactgtggga ggccgacgca ggaggattgc ttgaggccag gagttcaagg ccagcctgga 2400
caacatagta agacctatc tctaccccc taataaatta atttaaaaag ccccccaatc 2460
tgtggtatct tattatggca gccctagcaa gctaatacag tggtttgaga ggctgggagg 2520
gttgagggga agataaactt ttaaaaagct cttatctttc atttcaatca gttaaaaata 2580
```

```

cttgctcagt gtaacaattt tgcttctcag cttccactct aatattgttg tgccattaag 2640
caatttagct aatcctgaca tttcttagat tcataatgtt aggagcattt aatctgtatt 2700
ttacaagtta ggaagcagag gatcagagat gggaaaggac tagcccaagg ccaacattaa 2760
caagccctct aacaaaaact ttacaatata tttatgttga atggaactcc aagatctcac 2820
ctctccatcc aggaatggag tccatgtaat caaagtgaac ttaaaaatag gacagtttca 2880
acaagtcagg agattcacag caactgatca aagggagtcc agtcaacgtg agcaagcgtg 2940
attatgatga ggaagccccc tctgctttaa tccacacaag gaacgtaacc tgaagtaacc 3000
tgatgttaac caatctgctg tgtctactat gctgtttcct tgttcctgct agtgctgctt 3060
tacaaatgca gaccattcta tcatacctgg cggggcttct gttttatttt gtaggctgga 3120
tgctaccag ttcatgaatc gctaataaaa gcccaattaga tctttaaaaa aaaaaaaaaa 3180
aaaaaaaaa aaaaaaaaaa anaanaagan naannaaan anagaaga 3228

```

<210> 25

<211> 3106

<212> ADN

<213> Homo sapiens

<400> 25

```

accgaagtga gcgtggacct gccgaaacat gccaaaggcc aggatttggt tgatcagatt 60
gtgtaccact tggaccttgt ggaaacagat tactttggcc tccagttcct cgactctgcc 120
caggttgccg actggctgga tcatgcaaaa ccataaaaa agcagatgaa aattggacct 180
gcttatgctt tacactttcg agttaaatac tattcttcag aaccaaaaa ccttcgtgag 240
gagtttaciaa ggtacctgtt tgttttaciaa ctcaggcatg acattcttcc tggaaaaattg 300
aaatgccctt atgaaacagc tgtggaatta gctgctctct gtctacaagc ggagcttggg 360
gagtgcgagc ttccagaaca cacaccagag cttgtgtctg agtttcgggt cattccaaat 420
cagacagaag caatggaatt tgatatcttc cagagatgga aagagtgcag gggaaagagc 480
cctgcccagg cggaactctc ctatctgaat aaagcgaagt ggctggaaat gtatggggta 540
gacatgcacg ttgtcagggg aagagatggc tgtgaatatt ctcttggaat gaccccgaca 600
ggcatattaa tctttgaagg agctaacaaa ataggcttat tcttttggcc taaaattacc 660
aaaatggatt ttaaaaagag caaattgaca ctctgggtgg tctaggatga tgatcagggg 720
cgtgagcaag agcacacgtt tgtgttcagg ttagacagtg ccaggacctg caaacacctt 780
tggaagtgtg cagttgagca ccacgcattc ttccgactgc ggacgccagg aaacagcaaa 840
tccaatagat ccgactttat caggctgggc tctcgcttca gattcagtgg gcggacagaa 900
tatcaagcta cacatggctc caggttacga agaaccagca cctttgagag gaagcctagt 960
aaacgttatc catcccggag acattcaacg ttcaaagcaa gcaaccagat gatagcagcc 1020
cagctctgct ctaaaacaaa tccagaagtc cataattacc agcctcaata tcatcctaata 1080
atccatccca gccagccccg gtggcatcct cactctccaa atgtcaggcc atcctttcag 1140
gatgacaggt cgcattggaa agcatcggcc agtggagatg acagccattt tgattatgtc 1200
cacgaccaga accagaagaa cttaggaggg atgcaaagta tgatgtatcg agataaaactc 1260
atgactgcac tttgagagac tgaagcatct ctcttccatt caccttcata gtttcattgc 1320
attccatgaa aagtgtcttg gcctcagatg gatggatgtg tttggacgag tgtctttaag 1380
gagtagtcct gaaagggtgtt tttgggtgtc atgtaaatat ttgaagataa aaccactata 1440
gcttgtcata atttactgtt gactgcattc tcattaaaat gaaggtaaag gctcaggaat 1500
catattgatg tttctgattt aaaattggag tcaaagtcta tgtttatcat ttactatgt 1560
tctgatgtt ctttgttatt taattaatgg gagcaaataa aaccagaaga gcttgggaag 1620
attgctcagc atatattcct gtcgtagaag ttgagattgc taggggccag tttccctagt 1680
gtggcctgga cgagtcattt ccccttcatt gacctcattt tccccatctg aaaagagagg 1740

```

```

gttggactaa gtgatctcca aggtcctttc caactctaaa attctgcaat ttgttaacat 1800
ttcattttgt ttaggttgag gacatacatt caaactaatt ttatcacaa gaaaactgca 1860
ataccactt ccttgacaga gttactcctt tcagaagcta aataaagtat ataacttatt 1920
agatgttata tagatacagg gggactttga atttcacatc ttaaagcagt tgagctactt 1980
tgaatttaag cagtcgtact aatcttaaat tgcatagcat ttgttttgat cgaatttgct 2040
gctcaagtat ggggaataatt tttaatgtct taatgattgg tgctgctaac ttgcgtgatt 2100
tcagaagaca taattgtgaa tacacactgt cagaattggg ggattgggtt ttaccctaga 2160
cttcactctt aaaaagcaac gtgcaatcaa gatcatttat ggctcaaatg aaagcatata 2220
aggttttctt gaagttgtgc caaagcattc tgtagagtag gatgagatgg ttgttgccct 2280
agtctgttgg tagaaccaga aatcaatatg ttgtctttta gggttaaagct tgtacaaaa 2340
tattttatttc ccccatttca agccctgagt caaacatttt tttctcttaa taatagacct 2400
gaaatgtttt attagtattt ctgtgaaatc agttgattct tgtgccattt ttgtatatgt 2460
aattgtaatt ttgccatgt taggacctct aaaaaatgtt tgacatcctt tgagatattt 2520
tattactaaa atctgatctt ttttggtac tgcaaaaatc tattcagcaa gaaggtatca 2580
gctgcatacc ttgcacagtg gagctgacta cctataaact ctccctaagg catttgttta 2640
caggtgtatt ccattttagc agacgttctg atgctcagtg tatgtgctgc atacaaataa 2700
atgtgttctg aatcttttca tcttattgat agcattttta caaatgtgtt tccaaggaat 2760
aaagattatt cttgcttttt ttttgactcc atcttcattt tttttaaatt gattcttggt 2820
gctatgcaga agtctcattt gtgaatgacc ttggtaacag aacagttggc ttttggaagt 2880
ctgaaggtga gcattcagtt aggtgggtgg agcaagatca tcctagaatg aggtgctct 2940
tggaagagt ggatcttata ggcacagcag ctgatgcctt tctcatctg gggcaactct 3000
ggtgaaggtt gtccctgctg tcacaggtgc tgagtagaga gaagtgggtg cagtgggatt 3060
tcctcagtaa tagtctgtga aaggtacgtg tttgtcctgg ctactt 3106

```

<210> 26

<211> 3313

<212> ADN

<213> Homo sapiens

<400> 26

```

gggggccccct gctcaccggc ggcgcggccg tgcacatctc cgccgcgggc gccgccaaagg 60
ccaccctcta ctgccgcgtc ttcttgctcg acgggaccga agtgagcgtg gacctgccga 120
aacatgccaa aggccaggat ttgtttgatc agattgtgta ccacttgga cttgtggaaa 180
cagattactt tggcctccag ttctcgact ctgccaggt tgcgcactgg ctggatcatg 240
ccaaacccat aaaaaagcag atgaaaattg gacctgctta tgctttacac ttctgagtta 300
aatactattc ttcagaacca aacaaccttc gtgaggagtt tacaaggtac ctgtttgttt 360
tacaactcag gcatgacatt ctttctggaa aattgaaatg cccttatgaa acagctgtgg 420
aattagctgc tctctgtcta caagcggagc ttggggagtg cgagcttcca gaacacacac 480
cagagcttgt gtctgagttt cggttcattc caaatcagac agaagcaatg gaatttgata 540
tcttcagag atggaaagag tgcaggggaa agagccctgc ccaggcggaa ctctcctatc 600
tgaataaagc gaagtggctg gaaatgtatg gggtagacat gcacgttgct aggggaagag 660
atggctgtga atattctctt ggactgaccc cgacaggcat attaactctt gaaggagcta 720
acaaaatagg cttattcttt tggcctaaaa ttaccaaaat ggatttttaa aagagcaaat 780
tgacactcgt ggtggctcag gatgatgatc agggacgtga gcaagagcac acgtttgtgt 840
tccggttaga cagtgccagg acctgcaaac acctttggaa gtgtgcagtt gagcaccacg 900
cattcttcct actgcggacg ccaggaaaca gcaaatcaa tagatccgac tttatcaggc 960
tgggtctctc cttcagattc agtgggcgga cagaatatca agctacacat ggctccaggt 1020

```

```

tacgaagaac cagcaccttt gagaggaagc ctagtaaacy ttatccatcc cggagacatt 1080
caacgttcaa agcaagcaac ccagtgatag cagcccagct ctgctctaaa acaaattccag 1140
aagtcataaa ttaccagcct caatatcatc ctaatatcca tcccagccag ccccggtggc 1200
atcctcactc tccaaatgtc aggccatcct ttcaggatga caggtcgcat tggaaagcat 1260
cggccagtgg agatgacagc ctttttgatt atgtccacga ccagaaccag aagaacttag 1320
gagggatgca aagtatgatg tatcgagata aactcatgac tgcactttga gagactgaag 1380
catctctctt ccattcacct tcatagtttc attgcattcc atgaaaagtg tcttggectc 1440
agatggatgg atgtgtttgg acgagtgtct ttaaggagta gtctgaaag gtgtttttgg 1500
tgtccatgta aatatttgaa gataaaacca ctatagcttg tcataattta ctgttgactg 1560
cattctcatt aaaatgaagg taaaggctca ggaatcatat tgatgttctg attttaaaat 1620
tggagtcaaa gtctatgttt atcattttac tatgttcctg atgttctttg ttatttaatt 1680
aatgggagca aataaaacca gaagagcttg ggaagattgc tcagcatata ttctgtctgt 1740
agaagttagg attgctaggg tccagtttcc ctagtgtggc ctggacgagt catttccctt 1800
tcattgacct cattttcccc atctgaaaag agaggggttg actaagtgat ctccaaggct 1860
ctttccaact ctaaaattct gcaatttggt aacatttcat ttgttttagg ttgaggacat 1920
acattcaaac taattttatc acaaggaaaa ctgcaatacc cacttccttg acagagttac 1980
tcctttcaga agctaaataa agtatataac ttattagatg ttatatagat acagggggac 2040
tttgaatttc acatcttaaa gcagttgagc tactttgaat ttaagcagtc gtactaatct 2100
taaattgcat agcatttggt ttgatcgaat ttgctgctca agtatgggaa taatttttaa 2160
tgtcttaatg attggtgctg ctaacttgcg tgatttcaga agacataatt gtgaatacac 2220
actgtcagaa ttgggggatt ggtttttacc ctagacttca ctcttaaaaa gcaacgtgca 2280
atcaagatca tttatggctc aaatgaaagc atataagggt ttcttgaaat tgtgccaag 2340
cattctgtag agtaggatga gatgggtgtt gccctagtct gttggtagaa ccagaaatca 2400
atatgttgct ttttaggtta aagcttgtag caaaatattt atttcccca tttcaagccc 2460
tgagtcaaac atttttttct ctttaataa gacctgaaat gttttattag tatttctgtg 2520
aaatcagttg attcttgtag catttttgta tatgtaattg taattttgcc catgttaggc 2580
cctctaaaaa atgtttgaca tcttttgaga tattttatta ctaaaatctg atcttttttg 2640
gtactgcaa aaatctattc agcaagaagg tatcagctgc atacctgca cagtggagct 2700
gactacctat aaactctccc taaggcattt gtttacaggt gtattccatt ttagcagacg 2760
ttctgatget cagtgtatgt gctgcataca aataaatgtg ttctgaatct tttcatctta 2820
ttgatagcat ttttcaaat gtgtttccaa ggaataaaga ttattcttgc tttttntttt 2880
gactccatct tcattttttt taaattgatt ctgttgcta tgcagaagtc tcatttgtag 2940
atgaccttgg taacagaaca gttggctttt ggaagtctga aggtgagcat tcagttaggt 3000
gggtggagca agatcatcct agaatgaggc tgctcttggc aagagtggat cttataggca 3060
cagcagctga tgcctttctt catctggggc aactctggtg aaggttgtcc tgcctgtcac 3120
aggtgctgag tagagagaag tgggtggcagt gggatttcct cagtaatagt cctgtaaagg 3180
tacgtgtttg tcttggttac ttgtgctctt cctggcagga aggcattcaa acccttatct 3240
gtgggctcct ggaaattgtg tatgccatat aataccctct aataaatacc tctctgctta 3300
aaaaaaaaaa aaa 3313

```

<210> 27

<211> 1539

<212> ADN

<213> Homo sapiens

<400> 27

ctgaggatga ttatggtgat gacccttcga ccaactcgtt tgagaaagag aagcaggact 60

```

atgtctattg cctggagtc tccctgcaga cctacaaccc agactacgtc ctgatggtag 120
aagacgatgc tgtaccagaa gagcagatct tcccagtcct ggagcacctt ctgcgggctc 180
gcttctctga gccacatctc agagatgccc tttatctcaa gctgtatcac cccgagagge 240
tccagcacta catcaatcca gagcccatgc ggatcctgga atgggttggt gtaggcatgt 300
tgctggggcc cttactaacc tggatataca tgaggtttgc cagccgcccc gggtttagct 360
ggcctgtaat gctcttcttc tccctgtata gcatgggtct ggtggagctg gtgggtcggc 420
actatttccct ggaactgcgg cggtgagtc cttccctgta cagtgtggtt cctgcctctc 480
agtgttgac cccagccatg ctcttcccgg cactgcggc ccgcccggac ctcacctacc 540
tgtcccaagt gtactgccac aagggttttg gcaaggacat ggcaactgtac tcgctgttga 600
gggccaaggg agagagggcc tatgtagtgg agccgaacct cgtgaaacac atcgggctct 660
tctccagtc ccggtacaac tttcatccca gtctcctcta gggtgccaag agatgccttt 720
ctgaagttgg ccacttcttg aagattcaaa ttttatctc tttatttaga catggttgcc 780
tgcaggtatt tcaactgtta ctgttgtag agatattaggc actggggcag ctgaggaacc 840
tcaatatgtt aagagccttg gctttgtag cctcctggca ggagcagcag ttgcccacag 900
gtccggacct ctccctccac acagccacac tgcctcatgc agtctgacct acccagttag 960
ggtgcatttg aacactgatt atattctcca tttgttttta agctctgctt tgtgttagag 1020
cttgtgactg ccaaaaattt tgtgcacagt gatatgactg ttttaggatc ttaagggtag 1080
aattttgtga aagggtgagat cctttggaat tgagttcttt ctcatgggtt atgaaaatgg 1140
atgtatgttt agaatatatg cccaacgagg caggaccatg tggatagatt ccatttgttt 1200
ccttgacctg atgtaataaa aactgataaa agccgtgcag tgcccggcat cttggaggca 1260
ggtgatctgg cttttctgtt acctgggatt ttcttttctt atatatctgt agtgcactga 1320
taattgatat tgttagtgag aactttaagg cacgacaaga tctggaagag cttgtgaacc 1380
atcaaagtca gattgcttat gttgagtagt gatttaacat tgaacatttc tgtactactc 1440
tgactctttg aaaattggac atttcagaat gatacaatat aggggagctc tttacaagta 1500
gcatatcatt tttcctgact ttccctggctg actgcttac 1539

```

<210> 28

<211> 3404

<212> ADN

<213> Homo sapiens

<400> 28

```

ctcttctctgt gcaacgtgga gcgtagtgtg agccattttg atgccaagtt gctctccaag 60
tatgtccctg tggccaatcg ctatgagggc actgaggatg attatggtga tgacccttcg 120
accaactcgt ttgagaaaaga gaagcaggac tatgtctatt gcctggagtc atccctgcag 180
acctacaacc cagactacgt cctgatggta gaagacgatg ctgtaccaga agagcagatc 240
ttcccagtc tggagcacct tctgcgggct cgcttctctg agccacatct cagagatgcc 300
ctttatctca agctgtatca ccccagaggg ctccagcact acatcaatcc agagcccatg 360
cggatcctgg aatgggttgg tgtaggcatg ttgctggggc cttactaac ctggatatac 420
atgaggtttg ccagccgccc agggtttagc tggcctgtaa tgctcttctt ctccctgtat 480
agcatgggtc tgggtggagct ggtgggtcgg cactatttcc tggaaactgcg gcggtgagt 540
ccttccctgt acagtgtggt tctgcctctc cagtgttgca ccccagccat gctcttcccg 600
gcacctgcgg ccgcccggac cctcacctac ctgtcccaag tgtactgcca caagggtttt 660
ggcaaggaca tggcactgta ctgcgtgttg agggccaagg gagagagggc ctatgtagtg 720
gagccgaacc tcgtgaaaca catcgggctc ttctccagtc tccggtacaa ctttcatccc 780
agtctcctct aggggtgcaa gagatgcctt tctgaagttg gccacttctt gaagattcaa 840
atatttatct ctttatttag acatgggtgc ctgcaggat ttcactgttt actgttgtag 900

```

```

gagatatagg cactggggca gctgaggaac ctcaatatgt taagagcctt ggcttttgga 960
gcctcctggc aggagcagca gtttgccaca ggtccggacc tctccctcca cacagccaca 1020
ctgcctcatg cagtctgacc caccagtgga ggggtgcattt gaacactgat tatattctcc 1080
atttgttttt aagctctgct ttgtgttaga gcttgtgact gccaaaaatt ttgtgcacag 1140
tgatatgact gttttaggat ctttaagggtg gaattttgtg aaagggtgaga tcctttggaa 1200
ttgagttctt tctcattggg tatgaaaatg gatgtatgtt tagaatatat gcccaacgag 1260
gcaggaccat gtggatagat tccatttggt tccttgacct gatgtaataa aaactgataa 1320
aagccgtgca gtgcccggca tcttgagggc aggtgatctg gcttttctgt tacctgggat 1380
tttcttttcc tatatatctg tagtgactg ataattgata ttgttagtga gaactttaag 1440
gcacgacaag atctggaaga gcttgtgaac catcaaagtc agattgctta tgttgagtag 1500
tgatttaaca ttgaacattt ctgtactact ctgactcttt gaaaattgga catttcagaa 1560
tgatacaata taggggagct ctttacaagt agcatatcat ttttctgac tttcctggct 1620
gactgcttac gaaaaaaaaa aaaaaaaaaa atggactgaa gtcagaagac ctgacctttt 1680
atttcagatc tctgccatgg cctaacgacg tgaccacctt gggcaggnc aatantcact 1740
ctgggnntca atttcttga accngtaaaa tgaggacagt aatacttaca tgttaaagtt 1800
gccttgagga ttaaatngca gaatatatgt aaaacacctt gcaagngctt agtccatagn 1860
agacacttgg caaacttttc tctctataa taaacatata caactcaagc ttnnngcgga 1920
aaaagatatt ggtgatgttg tttaaagtat gagatagggc atggagctgc cacctatgta 1980
ccagctattc cagatgacca agccctgcta gaataaagac tgacttttca ttgtgacctc 2040
tgaaggatgt gcagctcagc tttcacctat cccacgctt acccaagcca aaaagaagaa 2100
ttgttcattc aaactggaaa tgatctccag cttggaaatg agacttctca tgagcttatt 2160
ctaagatcag agtctgaaaa ttattttctc aaaagttaca aataaatgtc ttttaacagt 2220
actgtcagac tagtgaccaa caagttactt tatgaggatg gatgatgata ttagagttca 2280
agcaattgtc cagtttgttt cgatgccaca gtttgccttt aaattcccat agtgagcttc 2340
catcttgaga cccactcaac acagaaggat tacataaagg ttgatcccc acgctctaca 2400
gatgaacaaa ttgagaccca gagatgctaa actgctctgc aggtcacaaa cctgaacagc 2460
agcatagcca aaatgggaac ggagctcttc tgacttcaag cccctgcaact ttttcttaa 2520
ccacattaat tttatgaata tctatgtggt caaagtctct tttgtaaaat ctgagtataa 2580
gctgaactag atcatgtctc tgtaagctaa gatggcagaa acacattgtc atgttagttt 2640
ggaggagcaa acagatataa gtatgtgttg taatgttata gtgagctgtt atgtatcccc 2700
cagcaatgtc caaagtgtatc agagctacta gatgacttat tacaatgttc agatccact 2760
ccttcattcc tctctctca gtgacctgtc tacccecaag ctttagaact gagctctctc 2820
caacttagaa ctcatlagag tttgttgcat tagtgattgg atctcctcat gaaagttttt 2880
tgaccaatga cagtagagcc cctttaagct agccaggccc aagataggga agcactgctg 2940
gtttacattt ctgctctttc taattcagtt gtctattgca tgcagacatc ccctgccga 3000
cttctccatt atggaaggct aaggccatcc ttccttgctc agtttctcta cctacacact 3060
aacatgcctt tccccatgg agtgaactct gccagttgtc tttggacatt ttctgcagat 3120
gcaggctcag ttctgctctt ttgttcccac ccacttctag atcttctcc atttaatccc 3180
ttattcttat aaggagtcta atttttgtgt attgtccttg attagggcag ctgacagctt 3240
tacaccaaag ctgaagtctc tcaacttaate ctaagcaagg ggctaggag aactaagata 3300
tccttcccat atcaagggat ataggttact agaagggtgg aggtctctta acaggctcac 3360
tcacctttca gagtttaaat tatggtcagg cttaagaagc tctc 3404

```

<210> 29

<211> 792

<212> ADN

<213> Homo sapiens

<400> 29

```

ttgtacagca agttagctct tcagggagtt agctctgtct atcctgcccc tgtctgccta 60
gccactcctt ttctaggtag agtggacctt gctgttcacg tcatagatgg ccattccgtg 120
tacctagact cctttttatt cctttgccaa caagtaattg gaatgtttgg atttaagaac 180
atgtctgtgc tttgcaattt catgtgacca tggtgaaatt atgcacacat tctatcattg 240
tgattcaaca agtatttttt gaatgctgct gtgcatgagg catggtgtta gatggcagat 300
tgacagaaat gactaataca ggtatctggc ctcaaggagt ttacagtttc ataggagaaa 360
gaaaaacaaa actagcaatt ctggtaagca gtatgtgttt tggagagcac ttaacttcat 420
cagagagcat ggttgaagggt gtatctgtgt gtatgagaag gtgttggtga taattagaag 480
gcttccctaa agaggtcact cctgggggtt agccataaag atcaagtaga tacaagtggg 540
catgaagcta gtattttgct gccagtcatt tctgtagtta ggttgcctat gtggttggag 600
catgtatgta ttgttacagg tgtagtctgc ttggttccc taagagaatg cccagtcct 660
ggccagaata tacatgtctt ttttagggtt tgcaccaagg ggaatagaga gtgtctggct 720
cagtacaaat ctgtttccgt tactccatta tccagaaagc cactgaaatg gcattagcct 780
acacttaatg gg                                     792

```

<210> 30

<211> 2733

<212> ADN

<213> Homo sapiens

<400> 30

```

ccacttgcca tgaacatgcc acatgccagc aaagagaagg gaagaagatc tgtatttgca 60
actatggatt tgtagggaaac gggaggactc agtgtgttga taaaaatgag tgccagtttg 120
gagccactct tgtctgtggg aaccacacat cttgccacaa ccccccgagg ggcttctatt 180
gcatttgcct ggaaggatat cgagccacaa acaacaacaa gacattcatt cccaacgatg 240
gcaccttttg tacagacata gatgagtgtg aagtttctgg cctgtgcagg catggagggc 300
gatgcgtgaa cactcatggg agctttgaat gctactgtat ggatggatac ttgccaagga 360
atggacctga accttccac ccgaccaccg atgccacatc atgcacagaa atagactgtg 420
gtacccctcc tgaggttcca gatggctata tcataggaaa ttatacgtct agtctgggca 480
gccaggttcg ttatgcttgc agagaaggat tcttcagtgt tccagaagat acagtttcaa 540
gctgcacagg cctgggcaca tgggagtccc caaaattaca ttgccaagag atcaactgtg 600
gcaacctccc agaaatgcgg cagccatct tggtaggaaa tcacagctcc aggtctggcg 660
gtgtggctcg ctatgtctgt caagagggtt ttgagagccc tggaggaaag atcacttctg 720
tttgcacaga gaaaggcacc tggagagaaa gtactttaac atgcacagaa attctgacaa 780
agattaatga tgtatcactg tttaatgata cctgtgtgag atggcaaata aactcaagaa 840
gaataaacc ccaagatctca tatgtgatat ccataaaagg acaacggttg gacctatgg 900
aatcagttcg tgaggagaca gtcaacttga ccacagacag caggacccca gaagtgtgcc 960
tagccctgta cccaggcacc aactacaccg tgaacatctc cacagcacct cccaggcgct 1020
cgatgccagc cgtcatcggt ttccagacag ctggtagcag aaactcccac gggagaaata 1080
cacatataaa agaagttgat ctcttagaag atgatggaag tttcaatatt tcaatattta 1140
atgaaacttg tttgaaattg aacaggcggt ctaggaaagt tggatcagaa cacatgtacc 1200
aatttaccgt tctgggtcag aggtggtatc tggctaactt ttctcatgca acatcgttta 1260
acttcacaac gagggaaaca gtgcctgtag tgtgtttgga tctgtaccct acgactgatt 1320
atacggtgaa tgtgacctg ctgagatctc ctaagcggca ctcagtcaa ataacaatag 1380
caactcccc agcagtaaaa cagaccatca gtaacatttc aggatttaat gaaacctgct 1440

```

```

tgagatggag aagcatcaag acagctgata tggaggagat gtattttatc cacatttggg 1500
gccagagatg gtatcagaag gaatttgccc aggaaatgac ctttaatatc agtagcagca 1560
gccgagatcc cgaggtgtgc ttggacctac gtccgggtac caactacaat gtcagtctcc 1620
gggctctgtc ttcggaactt cctgtggtca tctccctgac aaccagata acagagcctc 1680
ccctcccga agtagaattt ttacggtgc acagaggacc tctaccacgc ctcagactga 1740
ggaaagcaa ggagaaaaat ggaccaatca gttcatatca ggtgttagtg cttcccctgg 1800
ccctccaaag cacattttct tgtgattctg aaggcgcttc ctcttcttt agcaacgcct 1860
ctgatgctga tggatacgtg gctgcagaac tactggccaa agatgttcca gatgatgcca 1920
tggagatacc tataggagac aggctgtact atggggaata ttataatgca cccttgaaaa 1980
gaggagtgga ttactgcatt atattacgaa tcacaagtga atggaataag gtgagaagac 2040
actcctgtgc agtttgggct cagggtgaaag attcgtcact catgctgctg cagatggcgg 2100
gtgttggact ggggtccctg gctgttgtga tcattctcac attcctctcc ttctcagcgg 2160
tgtgatggca gatggacact gagtggggag gatgcactgc tgctgggcag gtgttctggc 2220
agcttctcag gtgcccgcac agaggctccg tgtgacttcc gtccaggag catgtgggcc 2280
tgcaactttc tccattccca gctgggcccc attcctggat ttaagatggg ggctatccct 2340
gaggagtcac cataaggaga aaactcagga attctgagtc ttccctgcta caggaccagt 2400
tctgtgcaat gaacttgaga ctctgatgt acactgtgat attgaccgaa ggctacatac 2460
agatctgtga atcttggctg ggacttcctc tgagtgatgc ctgagggtca gctcctctag 2520
acattgactg caagagaatc tctgcaacct cctatataaa agcatttctg ttaattcatt 2580
cagaatccat tctttacaat atgcagttag atgggcttaa gtttgggcta gagtttgact 2640
ttatgaagga ggtcattgaa aaagagaaca gtgacgtagg caaatgtttc aagcacttta 2700
gaaacagtac ttttccctata attagttgat ata 2733

```

```

<210> 31
<211> 2694
<212> ADN
<213> Homo sapiens

```

```

<400> 31
ccacttgcca tgaacatgcc acatgccagc aaagagaagg gaagaagatc tgtatttgca 60
actatggatt ttagggaac gggaggactc agtgtgttga taaaaatgag tgccagtttg 120
gagccactct tgtctgtggg aaccacacat cttgccacaa ccccccggg ggcttctatt 180
gcatttgctt ggaaggatat cgagccacaa acaacaacaa gacattcatt cccaacgatg 240
gcaccttttg tacagacata gatgagtgtg aagtttcttg cctgtgcagg catggagggc 300
gatgctgtaa cactcatggg agctttgaat gctactgtat ggatggatac ttgccaagga 360
atggacctga acctttccac cggaccaccg atgccacatc atgcacagaa atagactgtg 420
gtaccctctc tgaggttcca gatggctata tcataggaaa ttatacgtct agtctgggca 480
gccagggttc ttatgcttgc agagaaggat tcttcagtgt tccagaagat acagtttcaa 540
gctgcacagg cctgggcaca tgggagtccc caaaattaca ttgccaagag atcaactgtg 600
gcaaccctcc agaaatgcgg cagccatctt tggtaggaaa tcacagctcc aggtggggcg 660
gtgtggctcg ctatgtctgt caagagggct ttgagagccc tggaggaaag atcacttctg 720
tttgcacaga gaaaggcacc tggagagaaa gtactttaac atgcacagaa attctgacaa 780
agattaatga tgtatcactg tttaatgata cctgtgtgag atggcaaata aactcaagaa 840
gaataaacc ccaagatctca tatgtgatat ccataaaagg acaacggttg gaccctatgg 900
aatcagttcg tgaggagaca gtcaacttga ccacagacag caggacccca gaagtgtgcc 960
tagccctgta cccaggcacc aactacaccg tgaacatctc cacagcacct cccaggcgct 1020
cgatgccagc cgtcatcggt ttccagacag ctgaagttga tctcttagaa gatgatggaa 1080

```



```

gtttcaatat ttcaatattt aatgaaactt gtttgaaatt gaacaggcgt tctaggaaaag 1140
ttggatcaga acacatgtac caatttaccg ttctgggtca gaggtggtat ctggctaact 1200
tttctcatgc aacatcggtt aacttcacaa cgagggaaca agtgcctgta gtgtgttttg 1260
atctgtaccc tacgactgat tatacgggtga atgtgaccct gctgagatct cctaagcggc 1320
actcagtga aataacaata gcaactcccc cagcagtaaa acagaccatc agtaacattt 1380
caggatttaa tgaaacctgc ttgagatgga gaagcatcaa gacagctgat atggaggaga 1440
tgtatttatt ccacatttgg ggccagagat ggtatcagaa ggaatttgcc caggaaatga 1500
cctttaatat cagtagcagc agccgagatc ccgagggtgtg cttggacctc cgtccgggta 1560
ccaactacaa tgtcagtcct cgggctctgt cttcggaact tctgtgggtc atctccctga 1620
caacccagat aacagagcct cccctcccg aagtagaatt tttacgggtg cacagaggac 1680
ctctaccacg cctcagactg aggaaagcca aggagaaaaa tggaccaatc agttcatatc 1740
aggtgttagt gcttcccttg gccctccaaa gcacattttc ttgtgattct gaaggcgctt 1800
cctccttctt tagcaacgcc tctgatgtg atggatacgt ggctgcagaa ctactggcca 1860
aagatgttcc agatgatgcc atggagatac ctataggaga caggctgtac tatggggaat 1920
attataatgc acccttgaag agagggagtg attactgcat tatattacga atcacaagtg 1980
aatggaataa ggtgagaaga cactcctgtg cagtttgggc tcaggtgaaa gattcgtcac 2040
tcatgctgct gcagatggcg ggtgttgga tgggttccct ggctgttgtg atcattctca 2100
cattcctctc cttctcagcg gtgtgatggc agatggacac tgagtgggga ggatgactg 2160
ctgctgggca ggtgttcttg cagcttctca ggtgccgca cagaggctcc gtgtgacttc 2220
cgtccaggga gcatgtgggc ctgcaacttt ctccattccc agctgggccc cattcctgga 2280
tttaagatgg tggctatccc tgaggagtca ccataaggag aaaactcagg aattctgagt 2340
cttccctgct acaggaccag ttctgtgcaa tgaacttgag actcctgatg taaactgtga 2400
tattgaccga aggtcacata cagatctgtg aatcttggct gggacttctc ctgagtgatg 2460
cctgagggtc agctcctcta gacattgact gcaagagaat ctctgcaacc tcctatataa 2520
aagcatttct gttaattcat tcagaatcca ttctttacaa tatgcagtga gatgggctta 2580
agtttgggct agagtttgac tttatgaagg aggtcattga aaaagagaac agtgacgtag 2640
gcaaatgttt caagcacttt agaaacagta cttttcctat aattagttga tata 2694

```

<210> 32

<211> 2765

<212> ADN

<213> Homo sapiens

<400> 32

```

cacagcagca cattcactat tcagggtgag ccacaggagg aaagaagaga aatagagttt 60
agacgtctgt gccacttgcc atgaacatgc cacatgccag caaagagaag ggaagaagat 120
ctgtatttgc aactatggat ttgtaggga cgggaggact cagtgtgttg ataaaaatga 180
gtgccagttt ggagccactc ttgtctgtgg gaaccacaca tcttgccaca acacccccgg 240
gggcttctat tgcatttgcc tgggaaggata tcgagccaca aacaacaaca agacattcat 300
tccaacgat ggcacctttt gtacagacat agatgagtg gaagtttctg gcctgtgcag 360
gcatggaggg cgatgcgtga aactcatgg gagctttgaa tgctactgta tggatggata 420
cttgccaagg aatggacctg aacctttcca cccgaccacc gatgccacat catgcacaga 480
aatagactgt ggtacccttc ctgagggtcc agatggctat atcataggaa attatacgtc 540
tagtctgggc agccagggtc gttatgcttg cagagaagga ttcttcagtg ttccagaaga 600
tacagtttca agctgcacag gcctgggcac atgggagtc ccaaaattac attgccaaga 660
gatcaactgt ggcaacctc cagaaatgcg gcacgccatc ttggtaggaa atcacagctc 720
caggctgggc ggtgtggctc gctatgtctg tcaagagggc tttagagacc ctggaggaaa 780

```

```

gatcacttct gtttgcacag agaaaggcac ctggagagaa agtacttta catgcacaga 840
aattctgaca aagattaatg atgtatcact gtttaatgat acctgtgtga gatggcaaat 900
aaactcaaga agaataaacc ccaagatctc atatgtgata tccataaaaag gacaacggtt 960
ggaccctatg gaatcagttc gtgaggagac agtcaacttg accacagaca gcaggacccc 1020
agaagtgtgc ctagccctgt acccaggcac caactacacc gtgaacatct ccacagcacc 1080
tcccaggcgc tcgatgccag ccgtcatcgg ttccagaca gctgaagttg atctcttaga 1140
agatgatgga agtttcaata ttccaatatt taatgaaact tgtttgaaat tgaacaggcg 1200
ttctaggaaa gttggatcag aacacatgta ccaatttacc gttctgggtc agaggtggta 1260
tctggctaac ttttctcatg caacatcggt taacttcaca acgaggggaa aagtgcctgt 1320
agtgtgtttg gatctgtacc ctacgactga ttatacgggt aatgtgaccc tgctgagatc 1380
tcctaagcgg cactcagtg aaataacaat agcaactccc ccagcagtaa aacagaccat 1440
cagtaacatt tcaggattta atgaaacctg cttgagatgg agaagcatca agacagctga 1500
tatggaggag atgtatttat tccacatttg gggccagaga tggatcaga aggaatttgc 1560
ccaggaaatg acctttaata tcagtagcag cagccgagat cccgaggtgt gcttggacct 1620
acgtccgggt accaactaca atgtcagttc ccgggtcttg tcttcgggaa ttcctgtggt 1680
catctccctg acaaccaga taacagagcc tcccctccc gaagtagaat tttttacggt 1740
gcacagagga cctctaccac gcctcagact gaggaagcc aaggagaaaa atggaccaat 1800
cagttcatat caggtgttag tgcttcccct ggccctccaa agcacatttt cttgtgattc 1860
tgaaggcgtt tcctccttct ttagcaacgc ctctgatgct gatggatacg tggctgcaga 1920
actactggcc aaagatgttc cagatgatgc catggagata cctataggag acaggctgta 1980
ctatggggaa tattataatg cacccttgaa aagagggagt gattactgca ttatattacg 2040
aatcacaagt gaatggaata aggtgagaag acactcctgt gcagtttggg ctcaggtgaa 2100
agattcgtca tcatgctgc tgcagatggc ggggtgttga ctgggttccc tggctgttgt 2160
gatcattctc acattcctct ccttctcagc ggtgtgatgg cagatggaca ctgagtgggg 2220
aggatgcact gctgtggggc aggtgttctg gcagcttctc aggtgcccgc acagaggctc 2280
cgtgtgactt ccgtccaggg agcatgtggg cctgcaactt tctccattcc cagctggggc 2340
ccattcctgg atttaagatg gtggctatcc ctgaggagtc accataagga gaaaactcag 2400
gaattctgag tcttcctgc tacaggacca gttctgtgca atgaacttga gactcctgat 2460
gtacactgtg atattgaccg aaggctacat acagatctgt gaatcttggc tgggacttcc 2520
tctgagtgat gcctgagggg cagctcctct agacattgac tgcaagagaa tctctgcaac 2580
ctcctatata aaagcatttc tgttaattca ttcagaatcc attctttaca atatgcagtg 2640
agatgggctt aagtttgggc tagagtttga ctttatgaag gaggtcattg aaaaagagaa 2700
cagtgcgta ggcaaatgtt tcaagcactt tagaaacagt acttttccta taattagttg 2760
atata 2765

```

<210> 33

<211> 4745

<212> ADN

<213> Homo sapiens

<400> 33

```

gaccagatcc atgactgttg gtggagcatt tcatactact gaagctgaag ctagtcttca 60
gagtttgact cagatatatg ccttacctga aattcctcaa gatcaaatg ctgcagaatc 120
gtgggaaacc ttagaagcgg acttaattga acttagccaa ctgggtcactg acttctctct 180
cctagtgaat tctcagcagg agaagattga cagcattgca gaccatgtca acagtgcctg 240
tgtgaatgtt gaagagggaa ccaaaaactt agggaaggct gcaaaataca agctggcagc 300
tctgcctgtg gcaggtgcac tcacggggg aatggtaggg ggtcctattg gcctccttgc 360

```

```

aggcttcaaa gtggcaggaa ttgcagctgc acttgggtggt ggggtgttgg gcttcacagg 420
tggaaaattg atacaaagaa agaaacagaa aatgatggag aagctcactt ccagctgtcc 480
agatcttccc agccaaactg acaagaaatg cagttaaaaa ccaaatttca gtattattgg 540
tgccaacatg tctatcctga ggacctttgc tgctgttggg cactccgtca ccttttggaa 600
cacaagtata tcaagatagt ggctactgat gttcaagtgg gattgaagtg tgataaatgg 660
atatattttg ttgtttgctg ggggtgttcat ggagatgtta agagattgag gccctgggct 720
gagggatatat aatgtatgtc aggtaaagt tgaagactgc caaggagcag attttctccc 780
tggaaatgtg aaaactgaac ctataactct gataaggact tgagatgtgt agaaacgttg 840
ggttatggaa gactagtttc ttccataacc ctgaattgga gaccttaatg ctaagtgtag 900
attattgagg tttgttagtg aggaaaagaa taagagttca gaagcctttg ttatcagata 960
gcgaaatcag ggcctagtga ggagcacagg tcgactacat aatggagtcc attggcgaac 1020
cctattgcaa tttgggtccaa ctatatcttc tggtagaagg aattaatgat gtaagaaaat 1080
gcaagaggct caacttctct tccaaaaatc ttctggcttc tgaactcttc ctctgcctct 1140
ctttaataaa ataacacaga atttcaagtg gtaggagact tattaagcca gtcaccaagc 1200
ttggtctgtc agcctgtctt ctaacacctc aaagatcttg tgccctgtgc tgtccctccc 1260
ttgtaattat gaaaagtctt ttggtttctg ggggtgaattc taccatgta taatgaggaa 1320
ttctctcata accttttttg tcttgtctgt catctctgtt catcccttc tataacctct 1380
aggtaaaaag aaaagaaaaa aagaaatttc gagatatttt caacattgtt agagtltggg 1440
ctaaaatgag caaggagaaa aaaaccacca agaacatttc ctggggcatg ttccagtttt 1500
gaggggtgat atatctgcca gatagggggg atctgaccca gtcttctttt cagctgggtct 1560
ctgggggggag ctgagaactc gcttgctacc tcacatcctt ttccccagac tttttatctc 1620
ctatgcatcc ctttgctttc tatagctggg gtttcttccc caaaatggcg ttcccatgct 1680
tacctttctc acattctaga caatgatgga caaagacgca tgcaagactc agaccggggg 1740
aatggtgtgg tgctaattct aacacctgac attcacagca agcatggccc agcccaaccg 1800
catgtctatc tcaaaccgca gaaaggcttt aatactggaa aaaaagaatt caagactaca 1860
ggcagctccc ctctgtaccc caactcattt aaaataggag gaatcacttt ttgccttact 1920
taacgctttt ttctgagcac agggatgggc acctgcaccc cagaagggtg gagctgtctc 1980
tctgccagga gctaaggttc attaggggat tggatgggtt atcacttctt tctttctgag 2040
tttactttta gtaactttta ttgatggcta ctttctatgt cctgtctaa agagactttc 2100
tctttcatac gtcttaaatc tcatcaatga aatccagtga aacagcacca tttcttagta 2160
tcattaaata actagaaagt atcaagtatt gctctctgct gctttatatc attaacatat 2220
taataatacc aagaaggaaa tactttgaat aagtgtcaga ttctgatcca gtattggaca 2280
cctgtgatat tggacacctg tgaggctggg ataattactt ttgaattaca cctcttctct 2340
agtttctgga ccttgctctg tcactttaac acagggtgat caaacctgaa tgaggatcag 2400
aactcaccca ggcacatact aaagcaagat tcctaaacct cagttccagg ggtaattctg 2460
acatcacccg tccagcatag tcagctgaaa ttataaatct aagaaacagt tacatcaaga 2520
ttctgctgtg tcatttaatt ctgaaactcc cagtattcta cccttcttca tcaactgcata 2580
ttaccccact cttccatccc aaattggcta tccttcagc ccaccaactt agcggcagca 2640
ctagggattc attataaggt aaatctgggt tacataaaga cctgaaggag gcctgtattt 2700
gaagctcaca cttggtattg gtatctctca tttttactga gccagtgtgg aataccactg 2760
tatgtactca tataagccct tgacttttac tgctcatcag gattggaata ttactctagc 2820
agtcttcaca cataggcaag ttacagtcct tttaaaaagn atctcatttc cctataatgg 2880
aacctaatag ccaacttttt catagaaatt gctagaagag tttgatcaac tataaatgat 2940
aaagtgttta taagcatagt cagtgtgaca cagaaaccaa tcttaaaatt gaatttaatg 3000
ttttatcata tcagattaaa tttttctccc atgtcttatt tttactgcaa caagttagaa 3060
agtgggaaca ctttgattaa tgtcttaaaa tttgtgggcc ctcatgttga taaaggcagc 3120
aatcctaagg accttttttt tttttaacat aatctgagaa tttctctgta gagcagagac 3180
tttcaaacct tttggctgta acccacagta aaaaacgcat ttatatcaaa ccttagaata 3240

```

```

tgtttaaatga acaataactta ccattctgat gctttttatt gtttcagttt ttaaaatatg 3300
ccagttgcaa cccactaaat tgatatctac caatgggttg caacccttag cttgaaaaaa 3360
acaccctcac agaggaactg gtatttcttg aataccttct gtttgccagg cacttcacca 3420
ggcattttac aagtaaggaa actgggcttc agagaaaata atttgagag gtttactcaa 3480
ctacaaaggg gtgaagccag gaatgttaac taggtctgtt gagctacaaa aacttttatg 3540
tctctcagac tatacagcct ctatacaaaa ttgagatggg gggtgggggc aggggctcat 3600
gcctgcaatc ccagcactta gggaggcaga ggccagagga tcacttgagc ccaggagttt 3660
gagaccagcc tgggcaacat agtgagactc tgtctgtatg aaaaaaatta agaattagct 3720
gggtgtggca tagcacacac ctgtgggtccc agctacatgg gaagctgaag tgggaggatc 3780
acttgaactc aggagcagcc ttggtgacag aacaagaccc tctctcaaaa aaatatataa 3840
aaaaagggtg gtcattccatt ctctcttacc aaacaggctt tgaaatgaca cattccattc 3900
atttgcatct ttttaaaaaa cttctgattc cttactgagt gtccagcagc ctcaaagttt 3960
ttaatggtag ctgatgcaga cataaacagt gctcaatttg gcccttaaac tataaaatca 4020
agaaagtgtt tttcaatccc atccacctgc ctgcaagatt tcttaatgtt cactagttat 4080
aaccattgtt taaacagtgc tttttgtgtt atttaaaaaa aaactttaat gcttttttaa 4140
acaaatztat cataattcat agatcaaata attatccttt aaaatgatac ccttgggaaa 4200
tcatgtactt actgtagtga tgctagtatt aatattactt agaccaatth tgaagctgct 4260
ctttcagaat tgcttccaaa gacattttgc agatcatccc agaaaagggg gtatgatggg 4320
gctgtgtaga actgaccaga gtctctggag gattttgagg ttatactgaa actgagtgtt 4380
gtacagggag aattgcatga gtccagaaac ttctctctgt gggctgcctg ccttctctgc 4440
ctcccttaag tgctctaaga tttttgtaca ggagtaagaa tcaaatactg gtaacatcaa 4500
tcacaagaag ttgaggaaac ctgtaataata gctagataat atacaacgtt tgtcttccat 4560
cagagtgcag aaaccaaacc atgctttgtg ttaaccttaa atatgaaagg tgtttctcag 4620
ggtecccttt gtccttcgtt gctgccatat gaaatcttac aagggaaggat gaggaaaagc 4680
ctgggggggag gttctctctg gaaatgaggt ggtttttttt gttattaagt agaaccgtgg 4740
ctgtg 4745

```

<210> 34

<211> 5241

<212> ADN

<213> Homo sapiens

<400> 34

```

tgcgccgtgc ccaccgcccg gcctcgagcg ccccgccggg aggtttttct atatgagtgg 60
agaagacagc tgttaccagg gaggtcatac aacatttttt taggatgtct gaagatgaag 120
aaaaagtga attacgccgt ctgtaaccag ctatccagaa attcattaag atagtaatcc 180
caacagacct ggaaagggtt agaaagcacc agataaatat tgagaagtat caaagggtga 240
gaatctggga caagtgtcat gaagagcata tcaatgcagg acgtacagtt cagcaactcc 300
gatccaatat ccgagaaatt gagaaacttt gtttgaaagt ccgaaaggat gacctagtac 360
ttctgaagag aatgatagat cctgttaaag aagaagcatc agcagcaaca gcagaatttc 420
tccaactcca tttggaatct gtagaagaac ttaagaagca atttaatgat gaagaaactt 480
tgctacagcc tcctttgacc agatccatga ctgttggtgg agcatttcat actactgaag 540
ctgaagctag ttctcagagt ttgactcaga tatatgcctt acctgaaatt cctcaagatc 600
aaaatgctgc agaatcgtgg gaaaccttag aagcggactt aattgaactt agccaactgg 660
tcaactgact ctctctccta gtgaattctc agcaggagaa gattgcagac attgcagacc 720
atgtcaacag tgctgctgtg aatgttgaag agggaaacaa aaacttaggg aaggctgcaa 780
aatacaagct ggcagctctg cctgtggcag gtgcactcat cgggggaatg gtagggggtc 840

```

```

ctattggcct ccttgcaggc ttcaaagtgg caggaattgc agctgcactt ggtgggtgggg 900
tgttgggctt cacagggtgga aaattgatac aaagaaagaa acagaaaatg atggagaagc 960
tcaattccag ctgtccagat cttcccagcc aaactgacaa gaaatgcagt taaaaaccaa 1020
atttcagtat tattggtgcc aacatgtcta tcctgaggac ctttgctgct gttggacact 1080
ccgtcacctt ttggaacaca agtatatcaa gatagtggct actgatgttc aagtgggatt 1140
gaagtgtgat aaatggatat attttgttgt ttgctggggg gttcatggag atgttaagag 1200
attgaggccc tgggctgagg gtatataatg tatgtcagg aaagtttgaa gactgccaaag 1260
gagcagattt tctccctgga aatgtgaaaa ctgaacctat aactctgata aggacttgag 1320
atgtgtagaa acgttgggtt atggaagact agtttcttcc ataaccctga attggagacc 1380
ttaatgctaa gtgtagatta ttgaggtttg ttagtggaga aaagaataag agttcagaag 1440
cctttgttat cagatagcga aatcaggggc tagtgaggag cacaggtcga ctacataatg 1500
gagtccattg gcgaacccta ttgcaatttg gtccaaactat atcttctggt gaaggaaatt 1560
aatgatgtaa gaaaaatgcaa gaggtcaac ttctcttcca aaaatcttct ggcttctgaa 1620
ctcttctctt gcctctcttt aaataaataa cacagaattt caagtggtag gagacttatt 1680
aagccagtca ccaagcttgg tctgtcagcc tgtcttctaa cacctcaaag atcttgtgcc 1740
ctgtgctgtc cctcccttgt aattatgaaa agttctttgg tttctggggg gaattctacc 1800
catgtataat gaggaattct ctcataacct tttttgtctt gtctgtcatc tctgttcate 1860
ccttcctata acctctaggt aaaaagaaaa gaaaaaaaga aatttcgaga tattttcaac 1920
attgttagag tttgggctaa aatgagcaag gagaaaaaaa ccaccaagaa catttccctgg 1980
ggcatgttcc agttttgagg ggtgatatat ctgccagata gggggtatct gaccagctct 2040
tcttttcagc tggctctctg ggggagctga gaactcgctt gctacctcac atccttttcc 2100
ccagactttt tatctcctat gcaccccttt gctttctata gctgggtgtt cttccccaaa 2160
atggcgcttc catgcttacc tttctcacat tctagacaat gatggacaaa gacgcattga 2220
agactcagac ccgggggaatg gtgtgggtgt aatctcaaca cctgacattc acagcaagca 2280
tgccccagcc caaccgcatg tctatctcaa accgcagaaa ggctttata ctggaaaaaa 2340
agaattcaag actacaggca gctccctctt gtaccccaac tcatttaaaa taggaggaat 2400
cactttttgc cttacttaac gcttttttct gagcacagg atgggcacct gcaccccaga 2460
aggtgtgagc tgtctctctg ccaggagcta aggttcatta ggggattgga tggtttatca 2520
cttctttctt tctgagttta cttttagtaa cttttattga tggctacctt tcatgtccct 2580
gtctaaagag actttctctt tcatacgtct taaatctcat caatgaaatc cagtgaaca 2640
gcaccatttc ttagtatcat taaataacta gaaagtatca agtattgtc tctgtgtctt 2700
tatatcatta acatattaat aataccaaga aggaaatact ttgaataagt gtcagattct 2760
gatccagtat tggacacctg tgatattgga cacctgtgag gctgggataa ttacttttga 2820
attacacctc ttctctagtt tctggacctt gctctgtcac tttaacacag ggtgatcaaa 2880
cctgaatgag gatcagaact caccagga catactaaag caagattcct aaacctcagt 2940
tccaggggta attctgacat caccgtcca gcatagtcag ctgaaattat aaatctaaga 3000
aacagttaca tcaagattct gctgtgtcat ttaattctga aactcccagt attctaccct 3060
tcttcacac tgcatattac cccactcttc catcccaaat tggctatcct ttcagcccac 3120
caacttagcg gcagcactag ggattcatta taaggtaa atctggtttaca taaagacctg 3180
aaggaggcct gtatttgaag ctacacttg gtattggtat ctctcatttt tactgagcca 3240
gtgtggaata ccactgtatg tactcatata agcccttgac ttttactgct catcaggatt 3300
ggaatattac tctagcagtc ttcacacata ggcaagttac agtcccttta aaaagnatct 3360
catttcctta taatggaacc taatagccaa ctttttcata gaaattgcta gaagagtttg 3420
atcaactata aatgataaag tgtttataag catagtcagt gtgacacaga aaccaatctt 3480
aaaattgaat ttaatgtttt atcatatcag attaaatatt ttctccatgt cttattttta 3540
ctgcaacaag ttagaaagtg ggaacacttt gattaatgtc ttaaaatttg tgggcctca 3600
tttgataaaa ggcagcaatc ctaaggactt ttttttttt taacataatc tgagaatttc 3660
tctgtagagc agagactttc aaaccttttg gctgtaaccc acagtaaaaa acgcatttat 3720

```

```

atcaaacctt agaatatgtt taatgaacaa tacttaccat tctgatgctt tttattgttt 3780
cagtttttaa aatatgccag ttgcaacca ctaaattgat atctaccaat gggttgcaac 3840
ccttagcttg aaaaaaacac cctcacagag gaactgggat ttcttgaata ccttctgttt 3900
gccaggcact tcaccaggca ttttacaagt aaggaaactg ggcttcagag aaaataattt 3960
gcagaggttt actcaactac aaaggggtga agccaggaat gttaactagg tctgttgagc 4020
tacaaaaact tttatgtctc tcagactata cagcctctat acaaaattga gatgggggtt 4080
gggggcaggg gctcatgcct gcaatcccag cacttaggga ggacagaggc agaggatcac 4140
ttgagcccag gagtttgaga ccagcctggg caacatagtg agactctgtc tgtatgaaaa 4200
aaattaagaa ttagctgggt gtggcatagc acacacctgt ggtcccagct acatgggaag 4260
ctgaagtggg aggatcactt gaactcagga gcagccttgg tgacagaaca agaccctctc 4320
tcaaaaaaat atttaaaaaa aggtgggtca tccattctcc tttaccaaac aggctttgaa 4380
atgacacatt ccattcattt gcattctttt aaaaaacttc tgattcctta ctgagtgtcc 4440
agcagcctca aagtttttaa tggtagctga tgcagacata aacagtgttc aatttggccc 4500
ttaaactata aaatcaagaa agtgtatttc aatcccatcc acctgcctgc aagatttctt 4560
aatgttcaact agttataacc attgtttaaa cagtgtcttt tgtgtaattt aaaaataaac 4620
tttaatgctt tttaaaacaa atttatcata attcatagat caaatgatta tcctttaaaa 4680
tgataccctt gggaaatcat gtacttactg tagtgatgct agtattaata ttacttagac 4740
caattttgaa gctgctcttt cagaattgcc tccaaagaca ttttgagat catcccagaa 4800
aagggggtat gatgggtgctg tgtagaactg accagagttc ctggaggatt ttgaggttat 4860
actgaaactg agtgcctgtac agggagaatt gcatgagtc agaaacttcc ttctgtgggc 4920
tgctgcctt cctgcctcc cttaaagtgt ctaagatttt tgtacaggag taagaatcaa 4980
atactggtaa catcaatcac aagaagtga ggaaacctgt aatatagcta gataatatac 5040
aacgtttgtc ttccatcaga gtgcagaaac caaacctgc tttgtgttaa ccttaaatat 5100
gaaagggtt tctcagggtc ccctttgtcc ttcgttgctg ccatatgaaa tcttacaagg 5160
aaggatgagg aaaagcctgg ggggaggttc tectcgaaa tgagggtggt tttttgttta 5220
ttaagtagaa ccgtggctgt g 5241

```

<210> 35

<211> 3014

<212> ADN

<213> Homo sapiens

<400> 35

```

cccactctcg ctgtttctct ccaactctctc tccccatgaa ggaagagacc actggagttt 60
gcatgtaccc tccaatcaaa acgaggctga taaaaacatt cccggttgat acaatgacct 120
cgtttctctga tactttcacc acagggccac agtttactgc agacttcaga gacagtaaat 180
tacaatgctg tcttgcccg acttccccgc tgatcccagc agcgaccctg aggcctttga 240
cagagaccgt ctccacagt cagaccattt acaccaccgc gaaacctgtt tctctggcag 300
ccagtgcaga gacactccgg caggaactgg agagagagaa gatgatgaaa agactgttga 360
tgaccgaact gtgaaattct ccccttgta cctggaagat ggcattgtgc cttctgtccg 420
tcttctttct tcgggctttg tgtgctcact ctacacagc atacaagtgt gtgctttgtt 480
cgcccaggtc tccatgggtt gtgaaagcca atttctggct tgacttttgt gggaaaagtt 540
attttatgtc tctaagcat tagagtttt ctattactct atgtagtga gacaggattt 600
gataagtcta ggaaaagaaa gatgggaaaa cgggattcct tttcagaagt acctgtgtgt 660
atctgttaat aaccacagg gttaatatga tgtaggatct tttactatca atttcaacca 720
tttgattttg tatgattgaa acttgcaccg agctttgact gtttgttaaa gattcatttt 780
taatgaaaga ataattcttt attgctgggt tttcatttac actgataaat acacagatct 840

```

```

tataaagtct ttaacattca tttgtattca gatgtgagta gaagaactaa aaaaagaaag 900
ttacatatca ctatgactga aggtacttca gcttaatctg aaatataatt taacttgtga 960
actccttgga tatgatatta tttggaataa acagaattta tcattgaacc caaagtagga 1020
aatgatagct tacattgtct aaaaatcctt acaagggttaa gatgattcaa tatcaagaat 1080
attcagaaaa ttatttctaa agttgatcga ttcattgtctg attgatagaa tcttgaccag 1140
aagaaatctt gctcttttta tatagtttca agaaatgtgt ttttaaattt ttattaatgc 1200
acttgaacaa ctttgcagga ataaagcaac cccctaacca caaaatatcc ctctaaatta 1260
gttccctagc tttctcaatg aatacacaca tatttttaca tagctatgat cgttgtgtac 1320
attctccttt gttttacttc tgggcctaac acttgtctcc tcttgtcaac acagattcta 1380
ctctcaccaa tttaaatgtc tttatatcca tgtaacatgg gtaacctcac ttcaccccat 1440
tattagatat ttgagttata tctaattttt cactcttata aatagtgtg ctatgaatgt 1500
ctgtaaaaaa aaaaaaactg ctcttctctt tggattattc ccttaggaat atctccaaag 1560
agggattaca aggtcaaaga gcatgaagta ttttatagct cttgttttat attgccagat 1620
tgctttctag aaagatccaa tctttgggtt ggaaggacct taaagggtcat ctagttagc 1680
ctccccaccc cctctgaatg cttgaatccc ctgcacaatt tatgatgcca ccagcaatgt 1740
ataagcattt ctgtttacca atagctctgc cagtattggg ttttgccatt tttatttatt 1800
tttgctagtt taatagggtat gtatagttgt tcttgaagag ttgttttatt tcattaattg 1860
ctagcaaggc tgagcacttt tccatgtgat gatttactag ttgtatttcc ttgtgtgtaa 1920
aatgttcatt catttcttat gaccacttgt taagaggaac tgatctcata tatttgtatc 1980
agaactgtat ttttatgtta tattgtatag tttgtctctc tgcccctctc cttaaaactg 2040
aatggtgcca ataatttgat actaatgact acaaaaaaag gtaatgcctc atttactagt 2100
attgttgtaa aatgaggaat gtatgtgaat attcagataa ccgaggatta accctttaag 2160
tgctgaatct ttaaaatttt aatatatttt tttttgaggg aaatctttct aaaatgtatt 2220
acgcacttcc ctgccttagt aaacagagta tactggagag tatttaacct tttcttgatg 2280
agtcatggtc atgattataa acatcagccc cttttatacc ttggtacggt gcagtgatat 2340
cattaagagc tatcaatatg tgtagggctt ggcttggcct tttataggat attatgctgt 2400
tctcactgat ggttttttac tgctctctgc tctgtcagtg gagctatccg gggcaattgt 2460
agcgtttggg tctttttacc cctatgtccc ccggtatac ttttaaaaca gcttttagctg 2520
ttctttatct tgtgcacatg atacaaaata tgttcccgtc caatatgggg ctgtcacttc 2580
ttgccaaccc agcacctct tctctttcta acctgcttcc tgaggcttct gctcttcacc 2640
tcctgctcgc tgatggaaac ctccagggca aagctgaagg tttcttgggg aagccaggaa 2700
agccagtatt tcctatgtgt cagatctgct tggcttccaa gaagggtatg atgggctttt 2760
tggccagtgt ttccaggagg ctctgggctt cctgcttctt ccccgcttcc cccagagttc 2820
acagatgttg aagtttctga aggttgacgt cactggaagt ctgaccacaa acaagttggc 2880
tgttactgta tttgaaaccc agtacctttg gcagctcacc tctaaccagt aaaataagag 2940
gattccatgg tttcagcagt tgtcctggtt atttgcctat ttaagtgaat gatgggtgct 3000
tggcattgtg actg 3014

```

<210> 36

<211> 1575

<212> ADN

<213> Homo sapiens

<400> 36

```

tccaggggag atccagaatg catgtgtctc tccccacgca tttatgtcat gttggtagct 60
ttagatcagc catgggtgaga aaagaacaaa agcttttagt tgtttttgtt ttgttttgga 120
gaatttgttt accagtaaat acatcactgc ctgtacccca aatgttacca gctccctgag 180

```

```

gtgtcccaca tactattgtg agttctcaga gcatgaactg tcttcagaag agcaggggcta 240
ggactttgtcc cagcatctgt gctcccatc caatcctctt tctcacagag aaccacttcc 300
catataagat gcttaaggct ctcaaaacag cagaacaatg aaacatactc tccctacact 360
tgcttagcca agagatacca ctcaggtaac ttttttcagg acatggaaga tctgtttcaa 420
ggagattttac tgctatttta tttggaagaa gctggcaact ggtcttgacc aaaatagaaa 480
aaaaaaaaaa agtccacaaa tttaatcact tgtagggaac ccatctatca aggtacccta 540
ccatatactt ttgtatttaa tagattactt agaaaccaca aaaataggaa tccttaccct 600
ttcaattcct gttcaaccct aaaaactgtg ataaacgctc ccaaccctgt ggtgatcagg 660
gttatgtaat gttcaaagat tcagacacac ctgggtttgg attcagttgc aactgggttg 720
ttatcacact cacttctttt tagctgtgtg agcatgaata atttacttag cctctctgtg 780
tctttccacc taattaaaga ggatcyacc cagggttatt ttgaggattc aaaaaggcat 840
gcaacacacc tggagcacia ttccactttc attcaactaa ttcccttccc tttccccttc 900
ttccccttct acaagatcaa tatgtaaagg agacatgagg cttactggtt gcttttgaac 960
acttacttag ttcttagcta caccactctt aaaattaact ggacattagt gtacagccca 1020
tgtccaagcc cagagagaaa acaatgggaa caatttcaag gtctcacca ctccttcatt 1080
tgcagagggg acaacagact ttctgacctg agaactggag aattttttaa acaaaatctc 1140
tcattccagc ccaacctatt taactttttg tggaggaatt ttacatggag gaagtgaagca 1200
catgtcatgc tagccaagag gacattattg tcattaaaga gaggcattat ttatacacc 1260
tgcaatgtgc acattaaaat atggaaattt taaaattatg accaagggtt tgaaacatat 1320
tggattacat gctcacattt aacaaagaga ggaaatgtgt ttcagtttct ggagtggttg 1380
gaatttacia gctaattgtt caataaatct actcaagata gttacctaag gctttgtggc 1440
aatgaccttg aactgagagc ctgtatctgg atttagcact tgaaagatct aactggatat 1500
ttgggtttaa agaatacat ttattcccaa atcggaatgc tttgttttct ctgtcagtta 1560
attgccagtt gccaa
1575

```

<210> 37

<211> 1607

<212> ADN

<213> Homo sapiens

<400> 37

```

gcgaatcacg acgggcggtg gccgttggtt cccggcgggg ggcggggttt ccgctccctg 60
gggcgcacgt cagtcaggag gcggaagcgc agcgggggcg ggaagggtgt agtgcccgca 120
gttgagctcc tcttgctaa gtggtcgcgc cccctttaag agcagcgatt gtaaggagag 180
gcgggtcccg tgcctcggg tcccaggta ttgtgaagtg ctgaccaatt gccactggac 240
atacttgaaa caaaatagga aaatggcagc aaactcttca ggacaagggt ttcaaaacaa 300
aaatagagtt gcaatcttgg cagaactgga caaagagaaa agaaaactac ttatgcagaa 360
ccagtcttca acaaatcatc ctggagctag cattgcactc tcgagacctt ctcttaataa 420
ggacttccgg gatcacgctg agcagcagca tattgcagcc caacagaagg cagnttttnc 480
gcatgctcat gcacattcat ctggatactt catcactcaa gactctgcat ttgggaacct 540
tattcttccc gttttacct gccttgacct agaataaga aaacatttgc gatggaaaag 600
tgactttgta atatcaaag ccaaagctac tatcattcag tgctacatga actgtgactt 660
taagaatttt ggtgaacttt gatatttttt gtttgtctga aagaaaggaa tgtgtaagt 720
aaagctgaaa gaagaataac caggatgatg agagctgtgg aagctgtatc gtccaaggaa 780
ttgattatgt accgtgactg taactttttt gtaatgctgt ttaactctca atcagactgt 840
gaactggatg gtcacgaagt cattcccaa ctcctagcaa gtttgactga atatatcatg 900
tccacagtag attttcaaga atcatttata gtacttaact ttaaagaaac aaggctgett 960

```



```

ttaaaaaatg aactaatagg cttaaatacaa ttgcatccat atttgctgtt tataggattg 1020
ctatcagtat accttttgcg tttatagtca acatgtatca tcttgaaata ttctttctgg 1080
acttataact acttccccct ttttcacttt aaaacaaacc tcaagaataa attactaacc 1140
agtcttaacc atcttttata aacatatgct cttataaatg ttgtgactag atgcaattaa 1200
aaataatagg gaatgtggta ggtttttaat ttgtacatcc tcttatttag tgttaccaca 1260
taaagtatga gtttgtgtgg ttctgttttc catttttgtt ctaactgaaa actttttggc 1320
tgggcacggg gcctcatgcc tgtaatccca gcactttggg aggccaaaggc gggcagatca 1380
cttgagatca ggagtttgag accagcctgg ccaacatggt gaaaccctgt ctctactaaa 1440
agtataaaaa attagccatg tgtggtggca cacgcctgta ntcccagcta ctcnngaggc 1500
tgaggcagga gaatcgcttg aacctgggag gcagggttg cagtgaactg agacggtgtc 1560
actgcactcc agcctgggtg acagtgaagc tttgtctcaa agaaaaa 1607

```

<210> 38

<211> 1161

<212> ADN

<213> Homo sapiens

<400> 38

```

agctcctctt gcctaagtgg tcgcgcccc tttaagagca gcgattgtaa ggagaggcgg 60
tcccgggtgc ctcggtgcc aggtgattgt gaagtgtga ccaattgcc ctggacatac 120
ttgaaacaaa ataggaaaat ggcagcaaac tcttcaggac aaggttttca aaacaaaaat 180
agagttgcaa tcttggcaga actggacaaa gagaaaagaa aactacttat gcagaaccag 240
tcttcaacaa atcatcctgg agctagcatt gcactctcga gacctctct taataaggac 300
ttccgggatc acgctgagca gcagcatatt gcagcccaac agaaggcagc tttgcagcat 360
gctcatgcac attcatctgg atacttcac actcaagact ctgcatttgg gaaccttatt 420
cttctgttt tacctcgct tgaccagaa tgaagaaaac atttgcgatg gaaaagtgc 480
tttgtaatat caaatgccaa agctactatc attcagtgt acatgaactg tgactttaag 540
aattttggtg aactttgata tttttgttt gtctgaaaga aaggaatgtg taagtgaag 600
ctgaaagaag aataaccagg atgatgagag ctgtggaagc tgtatcgtcc aaggaattga 660
ttatgtaccg tgactgtaac tttttgttaa tgctgtttaa ctctcaatca gactgtgaac 720
tggtgggtca cgaagtcatt ccccaactcc tagcaagttt gactgaatat atcatgtcca 780
cagtagattt tcaagaatca tttatagtac ttaactttaa agaaacaagg ctgcttttaa 840
aaaatgaact aataggctta aatcaattgc atccatattt gctgtttata ggattgctat 900
cagtatacct tttgcgttta tagtcaacat gtatcatcct gaaatattct ttctggactt 960
ataactactt ccccttttt cactttaaaa caaacctcaa gaataaatta ctaaccagtc 1020
ttaaccatct tttataaaca tatgctctta taaatgtgt gactagatgc aattaaaaat 1080
aatagggaat gtggtaggtt ttaatttgt acatcctctt atttagtgtt accacataaa 1140
atgatgagtt tgtgtggttc t

```

<210> 39

<211> 1332

<212> ADN

<213> Homo sapiens

<400> 39

```

gattaccggt gtganccacc angcccggcc aagaactgat tttaaaataa attccatctc 60

```

```

agcaataaca atgaaaaata aatgnatana caaatanatt tggctctcaag cattnagaaa 120
gaggtaattg taatgccctt gggctcganga atgtgnantc tatgtnaata gnttaattgg 180
cctttttgtt tggaaatctt cgcagcatag ctcatgcaca ttcatctgga tacttcatca 240
ctcaagactc tgcatttggg aaccttattc ttctgtttt acctcgcctt gaccagaaat 300
gaagaaaaca tttgcgatgg aaaagtgact ttgtaatatc aaatgccaaa gctactatca 360
ttcagtgtca catgaactgt gactttaaga attttgggtga actttgatat tttttgtttg 420
tctgaaagaa aggaatgtgt aagtgaaagc tgaaagaaga ataaccagga tgatgagagc 480
tgtggaagct gtatcgtcca aggaattgat tatgtaccgt gactgtaact tttttgtaat 540
gctgtttaac tctcaatcag actgtgaact ggatgggtcac gaagtcattc cccaactcct 600
agcaagtttg actgaatata tcatgtccac agtagatttt caagaatcat ttatagtact 660
taacttttaa gaaacaaggc tgccttttaa aaatgaacta ataggcttaa atcaattgca 720
tccatatttg ctgtttatag gattgctatc agtatacctt ttgcgtttat agtcaacatg 780
tatcatcctg aaatattctt tctggactta taactacttc cccctttttc actttaaaaac 840
aaacctcaag aataaattac taaccagtct taaccatctt ttataaacat atgctcttat 900
aaatgttgtg actagatgca attaaaaata ataggggaatg tggtaggttt ttaatttghta 960
catcctctta tttagtgtta ccacataaat gatgagtttg tgtggttctg ttttccattt 1020
ttgttctaac tgaaaacttt ttggctgggc acggtgcctc atgcctgtaa tcccagcact 1080
ttgggaggcc aaggcgggca gatcacttga gatcaggagt ttgagaccag cctggccaac 1140
atggtgaaac cctgtctcta ctaaaagtat aaaaaattag ccatgtgtgg tggcacacgc 1200
ctgtantccc agctactcng gaggtgagg caggagaatc gcttgaaact gggaggcagg 1260
ggttcagtg agctgagacg gtgtcactgc actccagcct gggtgacagt gagtctttgt 1320
ctcaaagaaa aa                                     1332

```

<210> 40

<211> 10419

<212> ADN

<213> Homo sapiens

<400> 40

```

accgntccg ggctccttcg gccncccat gggctgctgc agctccgct cctccgccgc 60
gcagagctcc aaacgagaat ggaagccgct ggaggaccgt agctgcacag acataccatg 120
gctgctgctc ttcatcctct tctgcattgg gatgggattt atttgtggct tttcaatagc 180
aacagggtgca gcagcaagac tagtgtcagg atacgacagc tatggaaata tctgtgggca 240
gaaaaataca aagttggaag caataccaaa cagtggcatg gaccacccc agcggaagta 300
tgtattcttt ttggatccat gcaacctgga cttgataaac cggaagatta agtctgtagc 360
actgtgtgta gcagcgtgtc caaggcaaga actgaaaact ctgagtgatg ttcagaagtt 420
tgcagagata aatggttcag ccctatgtag ctacanccta aagccttctg aatacactac 480
atctccaaaa tcttctgttc tctgcccaca actaccagt ccagcgagtg cacctattcc 540
attcttccat cgctgtgctc ctgtgaacat ttctgtctat gccaaagttg cagaggccct 600
gatcaccttt gtcagtgcac atagtgtctt acacaggctg attagtggag taatgaccag 660
caaagaaatt atattgggac tttgcttggt atcactagtt ctatccatga ttttgatggt 720
gataatcagg tatatatcaa gagtacttgt gtggatctta acgattctgg tcatactcgg 780
ttcacttgga ggcacagggt tactatggtg gctgtatgca aagcaaagaa ggtctcccaa 840
agaaactgtt actcctgagc agcttcagat agctgaagac aatcttcggg ccctcctcat 900
ttatgccatt tcagctacag tgttcacagt gatcttattc ctgataatgt tggttatgcy 960
caaacgtgtt gctcttacca tcgccttggt ccacgtagct ggcaaggctt tcattcactt 1020
gccactgcta gtcttccaac ccttctggac tttctttgct cttgtcttgt tttgggtgta 1080

```

ctggatcatg acacttcttt ttcttggcac taccggcagt cctgttcaga atgagcaagg 1140
ctttgtggag ttcaaaatct ctgggcctct gcagtacatg tgggtggtagc atgtgggtggg 1200
cctgatttgg atcagtgaat ttattctagc atgtcagcag atgacagtgg caggagctgt 1260
ggtaacatac tattttacta gggataaaaag gaatttgcca ttacaccta ttttggcatc 1320
agtaaactgc cttattcggt accacctagg tacggtggca aaaggatctt tcattatcac 1380
attagtcaaa attccgcgaa tgatccttat gtatattcac agtcagctca aaggaaaggga 1440
aaatgcttgt gcacgatgtg tgctgaaatc ttgcatttgt tgcctttggg gtcttgaaaa 1500
gtgcctaaat tatttaaadc agaatgcata cacagccaca gctatcaaca gcaccaactt 1560
ctgcacctca gcaaaggatg cctttgtcat tctggtggag aatgctttgc gagtggctac 1620
catcaacaca gtaggagatt ttatgttatt ccttggcaag gtgctgatag tctgcagcac 1680
aggtttagct gggattatgc tgctcaacta ccagcaggac tacacagtat ggggtgctgcc 1740
tctgatcatc gtctgctctt ttgctttcct agtcgctcat tgcttctgt ctatttatga 1800
aatggtagtg gatgtattat tcttgtgttt tgccattgat acaaaataca atgatgggag 1860
ccctggcaga gaattctata tggataaagt gctgatggag tttgtggaaa acagtaggaa 1920
agcaatgaaa gaagctggta agggaggccg tgcgtgattc cagagagcta aagccgatgg 1980
cttcgggagc aagttctgct tgaacctagc cgacggttat ggaaacccat tgacattcca 2040
aaacaatata tacacataac tatgtatttg tgtgtgtggg tgtgtgtata tatgtatatg 2100
tatgtgtgta tatatgtata tgtatataca cacacacaca taaatcagcc aaaatcagag 2160
aaaaggaaca gggatttaac accttttcta tgcttatttt tgtcaaacat gtactccttt 2220
catacgggtg gctttttaca ggcaacttcc gtcatttaac gttttcaact gtaattgtct 2280
taatggaaat gttaaaatcc atatctgatt aacattttta ataacttaga ggagatttta 2340
actttattta aaaataggta aaattattgt acctaatat gtctaaagtt tatcagggg 2400
taatttccct gatgtctgta taaaatcaag atcttatttt actgatgcat aagtcctagt 2460
gggtcaagac taggcatatg ctttcagata aataaggat tactccaatc agttttcccc 2520
aatcaaagaa gccatgtcat ttactttta gaaacataca attgggcccc atatgggaa 2580
tttcataata gttcatacat ttgtcagcca acattaaaag gtaaccaact cctcaggtat 2640
ttgtagttta ccctaacgct tctttaaaag aaagtaggta aaaaaagaaa agggtagata 2700
atctttcgta tgcaaacctt tcccttata tttgtctttc tttccttttt gacttttagta 2760
gcatcctcca cacatttgtg tgcttgattt gaaaggaagc tggggcacc cgcgagttta 2820
gcctttaagt ttctgtgtat tgatttgagc attaatgaat gctgggagga ataaagaagg 2880
gacagaaaca tggaacataa agcattgaaa attccggtgc ttgggcttcg gcttcagagt 2940
aacgtcagtg gcttaggggt aaacggccat ttatttcaaa tgcttgctat acaatctgaa 3000
aacacactgg cagggtgctc tctccttggc aattcattga gtatccagag ttctacgatg 3060
tttaactgaa gaattggcta atgttttgat cctccagtgt gactgttggt tttgtttggg 3120
gggtgggttg gggtttttgc tttttattcc tgaagcttac cagatatgaa tggctaatac 3180
tccattgttc tgcttggtgt aatgggtgaat gctttaagaa aaaaaagtgt aatttgctaa 3240
gaataattca tgatctgttt atgcgataac tctttttgt tacaattttt ttaaaaaaag 3300
ctatttttgt taatgtaaag taaatatttc agagcaaatt ttttaaactt attgcactaa 3360
atacaggctc tgtacaaaaa aaaaaaaaaa aaaaaaaagc ctcagcattt tatcattcca 3420
tggaaggaga atcttttgaa agaaagcatt gcctcctacc agaactagac agtgaattag 3480
atcggtatta tggaatgca tacaagtaat gtcactaggg cttaataagc agccgtttgc 3540
taatgtgctt cctttcaaag ggttggacct ttaaattgct gcaaaaggta aattgtattt 3600
ttttttaagt attggtgttc ttactctag ctaggctaaa atttgctaaa tgccttgggt 3660
tcttttaaaa gttcatgtaa tatttctgat ttttcagaat atttgcaata agagtctgga 3720
tttttaaaaa cacatgcata cacacaatta agagctcatg tcttagcaag atctgggaaa 3780
ccaacattgc gagagttagt attttgaaag aataattctc cagaagttaa catctaatat 3840
ctagtatcac caaacagtat cgctgttctc ttttattcat ttgaaatgaa tataattata 3900
taactaacia ttgtccaaat agatgagaga gcaaatcatg tgagaaaatt cagaatacca 3960

```

tctgtttcat agccgcacag attttggact ttcacaaaca ttgggaacta aatttagaat 4020
tggcaaaagt ctagaagatg ggtatcaaaa cagaagacat tccaggagct agcaatttta 4080
agaggtgtcc ctccaaagtg acctgatgga agtcctgaac ttggaaatta gggtctactc 4140
acttgacat ccctgcatca tggactgttg ctgctccctg ttccatattg cgcgaatctc 4200
agctatttgg aagctaccag gaatgctttc taattatcat ttgcaactag aactgtaatc 4260
agaaagaaat tttgtatttt tgtataactt gattgtgtgc cattttatat aacagggtcct 4320
gttttacaaa taaattttgt tttactaaaa aaaaaaaaaa aaaaaaaagg 4380
tggggggaaa agagttgcaa attaatggcc aggctgtttt ccttaagtaa attcatgtgc 4440
cgtgtagaaa tatgcagata tgcattacac aggcacacac agagagatat ttaatatgtg 4500
gaatatcact ctttcatgag ctttattcct tgattaatat gataaagcct cacagaggat 4560
tttaagcaat atttaaatgt gttgaggtta tgtttggata ttcctgctgc ctcttttcat 4620
tcatttcaag tcattcttca gctagctatg ggctgcctta ttgctattcg ctactgctt 4680
ccatcttctg cccaagttag aagaatagat gaagaacaga aatttctctg tgaaatgtgg 4740
atacccgaa atattcataa atcattgatt ctatgtggtg gtttttgtct tcttctgtgg 4800
tgaatcttca aaactgtatt caggactcat atttctagag ttttaattgg ggctctctat 4860
tgtcaattgt agtagtgacc agagtatcgt gggttttgcc atcagataat taaggctctg 4920
gtgcataaatt tagactttct aagctcctgc ctgaagaata aggtcttcca taatatggaa 4980
gagaaaagtt atatttcagt gtaaatccaa gagacccatc ttctatgaga ggcttaccag 5040
tgcattagta ataaaacatt agcaatttag cttaggcact gagattgtat aatccttgca 5100
tggatgagca gagctcaaag ccagtgttcc ctttaaagca tttaatgcaa tggctaattg 5160
ttaggattaa agtttttatt ttatcctata agataattaa gtcctaagtc actacaactt 5220
aagtagctta actgtatgtt ggtacccaaa ctttttctat aatggttaag gaggaggcac 5280
ttactgagtt tattttaaag ttatgctaga aatgttttcc tttttagag atgctctctc 5340
tctctctctc tttttttttt tttttttttt ttcctgagg gcattaggct gctgattgta 5400
agttatttcc aataccactg atcttgggtat ctgccaaggg cttccttgcc ttcctgggc 5460
agtttttaga tttcacccca gtagggggta tgaggcactg acccctttat tctggacca 5520
aactgcttta gagcagagtt aatacctctc cctcctctt caaaacagct acaggaactg 5580
caggcaccac atatgtgtga ggccacatca ctgcccctga ggcttcaact atttccacca 5640
tgcactatta ctagctacta acagccattt taagagatct cctagcctgc aaggtagaat 5700
tctgggatca gttccaaaac cagaactgcc agtaatggtg acttgaatat tttatttgat 5760
ttttggtcag taaccttgct atttaaatg gactttccaa tagccttaac atggcctctg 5820
agaagtttcc tccagaaagg tctttaagct ttcgcttttc aataagtaag ttctttactg 5880
cttttctact gctactttaa aaaaaaacac acacacaacc accaataaaa ctctcagagt 5940
ctggaggctt atcagtgcac ctgccacagt taatttctgt attcttttca tacattccat 6000
ctgacacata agagtgcact ataattggctt gcctagaact ggtaagaagt ggtctgtgaa 6060
tgtatattga aaaaatgtgg gttcataagt aaaagcgtaa cgcagatatt tgtgtattct 6120
gtattcacag cgtaggctgc cttttgcttt aaatgcatat tcatgtatct ttgtgtgcta 6180
agaatgcatg ttageccttt tgaagtagt gcactattca tacacagtag cttcttgtaa 6240
ttaactcatc tttgttaact tagtggcaca tagtccaaat ttttaaaaag caagaccctt 6300
gaatatgcca agagaaaatc taaagctagc aaacttaaaa aacaaaataa aaattacag 6360
tcgtgtacag ttatgaattt agaatagcac actttttcca ttcagatttc atacatttga 6420
gccaaattct tatactccat gttttaattt taaaaggata attttaatcc aagattttaa 6480
tctttgaaaa tatctttctt atagaaaact ttaatgcagt ttttaaactt actgatttct 6540
gtggaaaacc tttcttttct atagaaatac tgtagtgcc tttcttccca tcttgatttt 6600
gtacatgtaa agacaaatga tgattcagtt tcaatattgc atgaacaatt gccactttgt 6660
aaatttatatg gacagaatgt gttctaagga aattcgttag taaatttggt aaaaacatgt 6720
gagattgttc gagacctatt aggctattct tcagttttga tgctcagttt tacaacttaa 6780
atgattttct ccaggacacg gagctcagaa taataaagct tttattaatg gtctagttaa 6840

```

gatctagttg	aacatggaaa	cattgtaatt	tacaaatgtc	tcagaaattt	ctcactttta	6900
tttgctaaga	cctgaattta	atatttgata	ttcaaaaaca	agttattttg	aagagacaat	6960
gggtctcttt	gagcttaaga	aagctatgga	ctatctttcc	ttcaaagca	catcacatgt	7020
ctgtgaacac	tcaaaatgct	catagaattt	cagggccctc	agacggccag	cttccacccc	7080
tgtacccccct	caggggctaa	tgaacccaag	tcaaaaggct	gctctaaagt	tgttctgata	7140
atgtgttttg	aaatgaggaa	gtgattagga	ctccactaga	gatacttttg	agatgatgct	7200
taccaccttc	tgacctggaa	tcatttttat	aacttaaaga	ctttatgtaa	tttagtagca	7260
ggtaggggaa	tagaaaccac	actaatctga	aggaataaca	agatgatcaa	cactgaatct	7320
tcatgacagt	aagattttct	ttccttgagg	gtatgagggt	gtgtgtgtgt	gtgtgtgtgt	7380
gtgtgtgtgt	ttgtgtgtgt	gtgtgtgtgt	gtcctggcgg	ttatttgggg	gcaaggatcc	7440
agattatttg	cagtgaagcca	aaaaaaaaaa	atatccaaga	agaaataaat	agggaaatct	7500
tgtaattaat	ccatctacca	aaacacttga	ttctttttat	tgtagctcag	ctatgacatt	7560
atgacattat	gtggcttagt	gttatcagta	tactactgtc	ttaaaatata	tttgtcaaca	7620
tgtccagttc	caacaacagt	tgaagatgat	gataaattag	caaactcaag	ggtagtccat	7680
agttggcaaa	aaataaattt	ggtagaaagt	tggaggagca	gatgctgaag	gtagaaatgc	7740
aaaagatgtt	ttttcttctt	gcataaatca	cggaggtgtg	ttctgtttta	gtgttttcca	7800
gtcaccctact	agagcagctt	cagattcctc	tgccaataga	actccacccc	cagtccttta	7860
cctactctat	cttcttttagg	gaagcatgtg	gatttcttca	catctgcacc	ttactattag	7920
gtaaggggtgc	tgcattctaca	tggaaaggctc	cctggggccct	tgaaaaagca	ggcagaaact	7980
catcaagggg	tttgaagttc	ttgtgaaatt	tgatcccagt	ggctcatgac	ttatagtcag	8040
gcatcaaacc	atttcttact	ttttcaggct	gaagaaagt	gaaatataac	tggctgggtga	8100
agaaaggaga	aaagtcagcc	ccctaccccg	ccccacactc	ctagaaaagt	ttggggggtt	8160
tttttgtttg	tttttggtgt	ttttttgaga	cggagtctcg	ctctatcgcc	agcttagagt	8220
gcagtggctc	aatcttagct	cactgcaacc	tccacctccc	aggttcaagt	gattctcctg	8280
catcagcctc	ctgagtagct	gggaccacag	gcattgtcca	ccacacccag	ctaatttttg	8340
tatttttagt	agagacgggg	tttcaccacg	ttggccaggc	caatctcgaa	ctcctgacct	8400
caggtgatcc	acccgcctca	gcctcccaaa	gtgctgggat	tacaggcatg	agccaccgcg	8460
cccggcctag	aagagtttta	atgaaccttc	cacagtagta	aatgcagaga	ccactgggtg	8520
aaattcccat	gttgataga	aaagggcgta	agtcagctct	aagtatctaa	atgtgatatg	8580
cccttttgct	acagaagtgt	aagacttaaa	gggaatatc	tgaccttcgt	gtatgaatct	8640
gatccacca	tcctgtcagc	taggattata	atttgaactg	tcgtttcagg	agcatgtgtt	8700
aaatcataatg	agtaaaaaaa	aagttagacat	tgaaaagaag	acttgggaat	aattgggcag	8760
atagaatggg	ttccatggat	cattctagtt	gccactagaa	aatgtgagct	ccttcttcat	8820
ttaccatggt	gataatccgg	tggtagcttt	tttttttttt	ttgtaaattg	tattagatac	8880
cccacaggaa	tgtgacaata	ataggatagc	tttgggggct	ggtgatttga	aacagggact	8940
attaagtaga	ttttccccc	tcctctaggt	tcctgtagtc	tgtgctcaga	acttggtttt	9000
tggcccttat	tgtttttgcc	tattttgatt	ttcagagatg	atcacatggg	gacagttaac	9060
ttttcttctg	ctgtgttgcc	ttaatgctac	tagattgtgt	tgtgttgttg	gagttttctg	9120
acttcttccc	tataaaaaa	tactgagagc	tccataatga	aagaagttgt	tatactttct	9180
cagaatattc	tggaccactg	aatgcacttc	taatagagct	ttaatctaaa	gaagttagtt	9240
cagtgggtat	taactgattt	tattacagga	gaaaaaaaact	ttacaaaaaa	ggcaggggaga	9300
aaagtgtgaa	gggcatcaag	caaaatgaca	ggggcttcaa	aaaacaacca	aagacaaaac	9360
cctatcttct	gaagaccaa	ggccaactt	tacttactgg	ctggcacagc	ctttctgaac	9420
tccttgagtt	tagaatagag	ctcctagaat	aataaggcgg	ccaaatttta	aagatcagtc	9480
aatacagtag	ggacctgcta	ttgatctctc	aggcactgag	ttcttcacat	ccagtgtcaa	9540
gcccagccca	gcatatgggg	tgatatgagc	agaaaacaca	catcgggtgtg	tcttgatttc	9600
tcgcagctgt	gtaatgtggc	atgagaagta	tgttttggtg	ccacatattt	ctcaatctga	9660
tgcccttttg	tctttttttt	tttttgccat	ttgcatccta	tttcatagtg	ccaaaatgaa	9720

```

tttttgtatc ttgtcttgtc tttgtccatt ataaactgga ggatcacagt taagccttcc 9780
atgaattcat agtttggaa catttacctt accattattt tggatttttt tcttattaca 9840
gtgtcactac actgtattca tgtgggggaa caaacatgta ggtgcttgaa catgccccac 9900
agacacagtt gtagccctag tatttgtgac ggtcattatt gacacagtgc agacttttnc 9960
gatggagcat tatgtcttca gaggacttta aaaatatgta tattaagcaa ttaactttct 10020
ggagttggag ttatcaaate ttgctgggaa attaacttcc aagagctctg aattggatgg 10080
aattccatct ggcttcagag aacaatcagc ctatatgaaa cggagctttg aaatgttttt 10140
cttgtgcaga aatacgaact agaaagaaaa gtgctgatct aatgctaagt tttctctgtg 10200
tactgactca gttgccagaa ttataatgaa aactgtattt tagtctaaca aatgtataga 10260
atTTTTtTatg taataaataa aatTTtatag gaaagaaaaa aaaaaaaaac tgtgttTtaa 10320
cacttcccat gccaggggtg gtgacgataa aatcatgggc aacatcttcg tttcaaatta 10380
ggtcagtatg ctgcaagggt gccccagaat tcccagctg 10419

```

<210> 41

<211> 1813

<212> ADN

<213> Homo sapiens

<400> 41

```

tgtgtgtagc agcgtgtcca aggcaagaac tgaaaactct gagtgatgtt cagaagtttg 60
cagagataaa tggttcagcc ctatgtagct acaacctaaa gccttctgaa tacactacat 120
ctccaaaatc ttctgttctc tgccccaac taccagttcc agcgagtgc cctattccat 180
tcttccatcg ctgtgtctct gtgaacattt cctgctatgc caagtttgca gaggccctga 240
tcacctttgt cagtgacaat agtgtcttac acaggctgat tagtggagta atgaccagca 300
aagaaattat attgggactt tgcttgttat cactagtctt atccatgatt ttgatgggtga 360
taatcaggta tatatcaaga gtacttgtgt ggatcttaac gattctgggc atactcgggt 420
cacttggagg cacagggtga ctatgggtgc tgtatgcaaa gcaaagaagg tctcccaaag 480
aaactgttac tcctgagcag cttcagatag ctgaagacaa tcttcggggc ctctcattt 540
atgccatttc agctacagtg ttcacagtga tcttattcct gataatgttg gttatgcgca 600
aacgtgttgc tcttaccatc gecttgttcc acgtagctgg caaggctctt attcacttgc 660
cactgctagt ctcccaaccc ttctggactt tctttgctct tgtcttgttt tgggtgtact 720
ggatcatgac acttcttttt ctggcacta cggcagtc tgttcagaat gagcaaggct 780
ttgtggagtt caaaatttct gggcctctgc agtacatgtg gtggtaccat gtggtgggce 840
tgatttggat cagtgaattt attctagcat gtcagcagat gacagtggca ggagctgtgg 900
taacatacta ttttactagg gataaaagga atttgccatt tacacctatt ttggcatcag 960
taaatcgctt tattcggttac cacctaggtg cgggtggcaaa aggatctttc attatcacat 1020
tagtcaaaat tccgcgaatg atccttatgt atattcacag tcagctcaaa ggaaagga 1080
atgcttgtgc acgatgtgtg ctgaaatctt gcatttgttg cctttgggtg cttgaaaagt 1140
gcctaaatta tttaaatcag aatgcataca cagccacagc tatcaacagc accaacttct 1200
gcacctcagc aaaggatgcc tttgtcatte tgggtggagaa tgctttgcga gtggctacca 1260
tcaacacagt aggagatttt atgttattcc ttggcaagg gctgatagtc tgcagcacag 1320
gtttagctgg gattatgctg ctcaactacc agcaggacta cacagtatgg gtgctgcctc 1380
tgatcatcgt ctgcctcttt gctttcctag tcgctcattg ctctctgtct atttatgaaa 1440
tggtagtggg tgtattatc ttgtgttttg ccattgatac aaaatacaat gatgggagcc 1500
ctggcagaga attctatatg gataaagtgc tgatggagtt tgtggaaaac agtaggaaag 1560
caatgaaaga agctggtaag ggaggcgtcg ctgattccag agagctaaag ccgatgctga 1620
agaaaagggt actggtctca tgagccctga agaatgaact cagaggaggt tgtttacatg 1680

```

```

aggttctccc actcaccagc tgttgagagt ctgcgattat gaagagcagg atcttattac 1740
ttcaatgaaa gcâtgtâaca agtttctcaa accaccaaca gccâagtgga tttggtacag 1800
tgcggctgtc taa 1813

```

```

<210> 42
<211> 1377
<212> ADN
<213> Homo sapiens

```

```

<400> 42
ggggcccccc cncncncca taataggntt tttgggannc ctggtcacct aacctcnaaa 60
gntngcceng gggnatatct ttgaggcaan aaagnccgac attgatggaa ggtcccaata 120
tgttactccc tncaccctc ccttaccatt tgggaagcag tgnctaaacc ctgggaaaat 180
acaaagnagt ctctgccta cacattcatc aaactttctg tatggtacca tggtttcata 240
cacctgtaac ccaggatatg aacttctggg gaacctctgt ctgatctccc aggaagatgg 300
aacttggaat ggcagtgcac catcctgcat ttcaattgaa tgtgacttgc ctactgctcc 360
tgaaaatggc tttttgcgtt ttacagagac tagcatggga agtgctgtgc agtatagctg 420
taaacctgga cacattctag cgggctctga cttaaggctt tgtctagaga atagaaagtg 480
gagtgggtgcc tccccacgct gtgaagccat ttcatgcaaa aagccaaatc cagtcatgaa 540
tggatccatc aaaggaagca actacacata cctgagcagc ttgtactatg agtgtgaccc 600
cggatatgtg ctgaatggca ctgagaggag aacatgccag gatgacaaaa actgggatga 660
ggatgagccc atttgcattc ctgtggactg cagttcacc cagtcctcag ccaatggcca 720
ggtgagagga gacgagtaca cattccaaaa agagattgaa tacacttgca atgaagggtt 780
cttgcttgag ggagccagga gtcgggtttg tcttgccaat ggaagtggga gtggagccac 840
tcccgaactgt gtgcctgtca gatgtgccac cccgccacaa ctggccaatg gggtgacgga 900
aggcctggac tatggcttca tgaaggaagt aacattccac tgtcangagg gctacatctt 960
gcacgggtgct ccaaaactca cctgtcagtc agatggcaac tgggatgcag agattcctct 1020
ctgtaaacca gtcaactgtg gacctctga agatcttgcc catgggttcc ctaatggttt 1080
ttcctttatt catggggggc atatacagta tcagtgcctt cctgggtata agctccatgg 1140
aaattcatca agaagggtgcc tctccaatgg ctctggagt ggcagctcac ctctctgcct 1200
gccttgacaga tgttccacac cagtaattga atatggaact gtcaatggga cagattttga 1260
ctgtggaaag gcagcccgga ttcagtgtt caaaggcttc aagctcctag gactttctga 1320
aatcacctgt gaagcngatg gcnngttnag ntntgggttc cccactgtg aacacac 1377

```

```

<210> 43
<211> 452
<212> ADN
<213> Homo sapiens

```

```

<400> 43
acaatcttga gcttcaagtc agtggaaaga aatgtacaat ctaattgaac aaaaaaaaaat 60
ttatatattca aaacattcaa agtgtcagat tggttgcata aattaaaaga gattagagaa 120
gaggaaaatc agtggacttg gagtgatcag caaaggcttc aataagtagg tagaacttaa 180
tctgggactt aaatgtgaaa attgaccagc ttgggaaggg aaaaaggagg gtggtaccac 240
attaaggatg tggaagtcaa gcaggaacag caggaaatgg gataatgcta acggtatatt 300
gaaggccag cagtggctgc ttgcaggga gttggagtcc aatcnttttc cntacgtgcc 360

```

```

ntttttgggg tatecttccn gatttgctta agtttttntt ttnattaatt gctnngggttt 420
tttttttccc gcngttttgc cegtntcccc cg 452

```

<210> 44

<211> 562

<212> ADN

<213> Homo sapiens

<400> 44

```

gtgcacatct tcagaaaact gatcctaagc aaatatcata atgaaaaacc atagattggt 60
aatggactaa gagaccttta gggttcattag ctccaaaccc ctcattttac agaagagaga 120
attgtaagat agatagaaaa ctagtggcag aatcaagagt aaaactcagg tctgatgctt 180
ctcagatctt tccataagtt agcacttggt tcccatgcct cttagttatc tcaaagaatg 240
acccccatcc gtataaacat cacagataat gcccagggtat ctgggacttt caatectggt 300
ttcatcaact cctaaagaga tgtagctctc tcagttcaat tttttctttt gaatctattt 360
ggttgctttt tttttttttt ccagggaaaa ggaggactgg gttttaacca ctgcaçgacc 420
atctgggctc tcccaaaagg cagggatcca tctctcctcg gtagtggcct ggggcaccc 480
ggggaactta tggcaaagga aagtcccaac atgggtgctg gggtcttgtt tagtnaaact 540
tgttnacttg ggggttactt tt 562

```

<210> 45

<211> 1766

<212> ADN

<213> Homo sapiens

<400> 45

```

anggaaaggg ttaattcccc nnttggaat tncnnggatt tccccagant tncaatttcc 60
accnttggtc aanaaaaaann ttgggnncc nctnnnttcn ccngtcanaa agannnnncn 120
cccttgaaag gnnccntnag agaattctct gcagtcnaa aaaatgtcct ctcccggaaa 180
acataacaca tatacttggt catggggacg atttcagtgt gaataggcaa gtttctgtgt 240
catgtgcaga aggggtatacc tttgaggag ttaacatata agtatgtcag cttgatggaa 300
cctgggagcc accattctcc gatgaatctt gcagtccagt ttcttgtagg aaacctgaaa 360
gtccagaaca tggatttggt gttggcagta aatacacctt tgaaagcaca attatttata 420
agtgtgagcc tggctatgaa ctagagggga acagggaacg tgtctgccag gagaacagac 480
agtggagtgg aggggtggca atatgcaaag agaccagggt tgaaactcca cttgaatttc 540
tcaatgggaa agctgacatt gaaaacagga cgactggacc caacgtggta tattcctgca 600
acagaggcta cagtcttgaa gggccatctg aggcacactg cacagaaaat ggaacctgga 660
gccaccagc cctctctctg aaaccaaata catgccctgt tccttttggt attcccagaa 720
atgctctgct gtctgaaaag gagttttatg ttgatcagaa tgtgtccatc aaatgtaggg 780
aaggttttct gctgcagggc cacggcatca ttacctgcaa ccccgacgag acgtggacac 840
agacaagcgc caaatgtgaa aaaatctcat gtggtccacc agctcacgta gaaaatgcaa 900
ttgctcgagg cgtacattat caatatggag acatgatcac ctactcatgt tacagtggat 960
acatgttgga gggtttctct aggagtgtt gtttagaaaa tggaacatgg acatcacctc 1020
ctatttgtag agctgtctgt cgatttccat gtcagaatgg gggcatctgc caacgccccaa 1080
atgcttggtc ctgtccagag ggctggatgg ggcgcctctg tgaagaacca atctgcattc 1140
ttccctgtct gaacggaggt cgctgtgtgg ccccttacca gtgtgactgc ccgccttggc 1200

```



```

tggacgggggt ctcgctgtca tacagctggt tggcagctct cctgctttaa tgggtggaaaa 1260
tgtgtaagac caaacccgat tcactgtcct tcttcttggg cgggacataa ctgttccagg 1320
aaaaggagga ctgggtttta accactgcac gaccatctgg ctctcccaa agcaggatca 1380
tctctcctcg gtatgcctg ggcatacctg aacttatgca aagaaagtcc aacatggtgc 1440
tggttcttgt ttagtaaact tgttacttgg gggtactttt ttattttgt gatataattt 1500
gttattectt gtgacatact ttcttacatg ttccatttt taaatatgcc tgtattttct 1560
atataaaaat tatattaaat agatgctgct ctaccctcac aaaatgtaca tattctgctg 1620
tctattggga aagtccctgg tacacatttt tattcagtta cttaaaatga tttttccatt 1680
aaagtatatt ttgctactaa ataatatnnn gctggatagn accattttat gagngnccaa 1740
gggatcatag agaggctcat ganttc 1766

```

<210> 46

<211> 601

<212> ADN

<213> Homo sapiens

<400> 46

```

cagtggaaag aatgtacaa tctaattgaa cacaaaaaaa tttatagttc aaaacattca 60
aagtgtcaga ttggttgcac aaattaaaag agattagaga agaggaaaat cagtggactt 120
ggagtgatca gcaaggctt caataagtag gtagaactta atctgggact taaatgtgaa 180
aattgaccag cttggaagga aaaaggaggg tggtaggaca taaggatgtg aagtcaagca 240
gaacagcaga aatggataat gctaacggta tatgaagacc agcagtggct gctgcaggaa 300
gtagagtcaa atcttatcct acgtgccttt atggatatca tcagaattgc taagtattaa 360
ttaaataaat gcatgggttt tttattccag ctgtttgcca gtctccctgc ttaaattggtg 420
gaaaatgtgt aagaccaaaac cgatgtcact gtctttcttc ttggacggga cataactgtt 480
ccaggaaaag gaggactggg ttttaaccac tgcacgacca tctggctctc ccaaaagcag 540
gatcatctct cctcggtagt gcctgggcat cctggaactt atgcaaagaa agtccaacat 600
g 601

```

<210> 47

<211> 3706

<212> ADN

<213> Homo sapiens

<400> 47

```

gaatggcaaa ttctcttaca cggacctaca ctatggacag accgttacct actcttgcaa 60
ccgaggcttt cggcncgaag gtcccagtc cttgacctgt ttagagacag gtgattggga 120
tgtagatgcc ccatcttgca atgccatcca ctgtgattcc ccacaacca ttgaaaatgg 180
ttttgtagaa ggtgcagatt acagctatgg tgccataatc atctacagtt gcttccctgg 240
gtttcagggtg gctggctcat ccatgcagac ctgtgaagag tcaggatggc caagttccat 300
cccaacatgt atgccaatag actgtggcct ccctctcat atagattttg gagactgtac 360
taaactcaaa gatgaccagg gatattttga gcaagaagac gacatgatgg aagttccata 420
tgtgactcct caccctcctt atcatttggg agcagtggct aaaacctggg aaaatacaaa 480
ggagtctcct gctacacatt catcaaactt tctgtatgg accatgggtt catacacctg 540
taanccagga tatgaacttc tggggaaccc tgtgctgac tnccaggaag atggaacttg 600
gaatggcagt gcaccatcct gcatttcaat tgaatgtgac ttgcctactg ctctgaaaa 660

```

tggctttttg	cgtttttacag	agactagcat	gggaagtgct	gtgcagtata	gctgtaaacc	720
tggacacatt	ctagcgggct	ctgacttaag	gctttgtcta	gagaatagaa	agtggagtgg	780
tgcctcccca	cgctgtgaag	ccatttcata	caaaaagcca	aatccagtca	tgaatggatc	840
catcaaagga	agcaactaca	catacctgag	cacgttgtac	tatgagtgtg	accccgata	900
tgtgtgaat	ggcactgaga	ggagaacatg	ccaggatgac	aaaaactggg	atgaggatga	960
gcccatttgc	attcctgtgg	actgcagttc	acccccagtc	tcagccaatg	gccaggtgag	1020
aggagacgag	tacacattcc	aaaaagagat	tgaatacact	tgcaatgaag	ggttcttctg	1080
tgagggagcc	aggagtccgg	tttgtcttgc	caatggaagt	tggagtggag	ccactcccga	1140
ctgtgtgcct	gtcagatgtg	ccaccccgcc	acaactggcc	aatggggtga	cggaaggcct	1200
ggactatggc	ttcatgaagg	aagtaacatt	ccactgtcat	gagggctaca	tcttgacagg	1260
tgctccaaaa	ctcacctgtc	agtcagatgg	caactgggat	gcagagattc	ctctctgtaa	1320
accagtcaac	tgtggacctc	ctgaagatct	tgcccattgt	ttccctaata	gtttttcctt	1380
tattcatggg	ggccatatac	agtatcagtg	ctttcctggg	tataagctcc	atggaaattc	1440
atcaagaagg	tgctcttcca	atggctcctg	gagtggcagc	tcaccttctc	gcctgccttg	1500
cagatgttcc	acaccagtaa	ttgaatatgg	aactgtcaat	gggacagatt	ttgactgtgg	1560
aaaggcagcc	cggatttcagt	gcttcaaagg	cttcaagctc	ctaggacttt	ctgaaatcac	1620
ctgtgaagcc	gatggccagt	ggagctctgg	gttcccccc	tgtgaacaca	cttcttggg	1680
ttctcttcca	atgataccaa	atgctgtcat	cagtgtgagc	agctcttgga	aggaaaatgt	1740
gataaacttac	agctgcaggt	ctggatatgt	catacaaggc	agttcagatc	tgatttgtac	1800
agagaaaggg	gtatggagcc	agccttatcc	agtctgtgag	cccttgctct	gtgggtcccc	1860
accgtctgtc	gccaatgcag	tggaactggg	agaggcacac	acctatgaaa	gtgaagtga	1920
actcagatgt	ctggaagggt	atacgtatga	tacagataca	gatacattca	cctgtcagaa	1980
agatgggtgc	tggttccctg	agagaatctc	ctgcagtcct	aaaaaatgtc	ctctcccga	2040
aaacataaca	catatacttg	tncatgggga	cgatttcagt	gtgaataggc	aagtttctgt	2100
gtcatgtgca	gaagggtata	cctttgaggg	agttaacata	tcagtatgtc	agcttgatgg	2160
aacctgggag	ccaccattct	ccgatgaatc	ttgcagtcca	gtttcttgtg	ggaaacctga	2220
aagtccagaa	catggatttg	tggttggcag	taaatacacc	tttgaaagca	caattattta	2280
tcagtgtgag	cctggctatg	aactagaggg	gaacagggaa	cgtgtctgcc	aggagaacag	2340
acagtggagt	ggaggggtgg	caatatgcaa	agagaccagg	tgtgaaactc	cacttgaatt	2400
tctcaatggg	aaagctgaca	ttgaaaacag	gacgactgga	cccaacgtgg	tatattcctg	2460
caacagaggc	tacagtcttg	aaggggccatc	tgaggcacac	tgacacagaa	atggaacctg	2520
gagccaccca	gtccctctct	gcaaaccaaa	tccatgccct	gttccttttg	tgattcccga	2580
gaatgctctg	ctgtctgaaa	aggagtttta	tgttgatcag	aatgtgtcca	tcaaatgtag	2640
ggaaggtttt	ctgctgcagg	gccacggcat	cattacctgc	aaccccgacg	agacgtggac	2700
acagacaagc	gccaatgtg	aaaaaatctc	atgtggtcca	ccagctcacg	tagaaaatgc	2760
aattgctcga	ggcgtaacatt	atcaatatgg	agacatgac	acctactcat	gttacagtgg	2820
atacatgttg	gaggggttcc	tgaggagtgt	ttgtttagaa	aatggaacat	ggacatcacc	2880
tcttatattgc	agagctgtct	gtcgatttcc	atgtcagaat	gggggcatct	gccaacgccc	2940
aaatgcttgt	tcctgtccag	agggtgggat	ggggcgcttc	tgtgaagaac	caatctgcat	3000
tcttccctgt	ctgaacggag	gtcgctgtgt	ggcccttac	cagtgtgact	gcccgcctgg	3060
ctggacgggg	tctcgctgtc	atacagctgt	ttgccagtct	ccctgcttaa	atgggtggaaa	3120
atgtgtaaga	ccaaaccgat	gtcactgtct	ttcttcttgg	acgggacata	actgttccag	3180
gaaaaggagg	actgggtttt	aaccactgca	cgaccatctg	gctctcccaa	aagcaggatc	3240
atctctcttc	ggtagtgcct	gggcatcctg	gaacttatgc	aaagaaagtc	caacatgggtg	3300
ctgggtcttg	tttagtaaac	ttgttacttg	gggttacttt	ttttattttg	tgatatattt	3360
tgttattcct	tgtgacatac	tttcttacct	gtttccattt	ttaaatatgc	ctgtattttc	3420
tatataaaaa	ttatattaaa	tagatgctgc	tctaccctca	caaaatgtac	atattctgct	3480
gtctattggg	aaagtctctg	gtacacattt	ttattcagtt	acttaaaatg	atttttccat	3540

```

taaagtatat ttgctacta aataanannn tgctggatag taccatttta tgaggtggcc 3600
aagggattca tagagaggac tcatgatctt tcatctgtgc cactggcacc acataggaga 3660
ccccttacia taaaggagac ccctttgttc tggttccatc ttctat 3706

```

```

<210> 48
<211> 373
<212> ADN
<213> Homo sapiens

```

```

<400> 48
tgagacattt ccactctggc attctagagt tctctggctt gcccttggtg ctttaaagct 60
ggttccaact tcgtgggctt tatgattgag caacactggg gattgtgatt tctaatacca 120
ccctccagat tcatgtcaca ggtgaaagtc tgcctgtggg gagagcttcc acattctcct 180
ggagttgaac caaataaagg aaaacatcaa caaacacagt taccaaaagc ttccaaggag 240
ggatggagga gaaattctgt tcattgcaga tggggagaaa ttaaatgaca tcaagtgcac 300
aatggaactg ggtaggattc aaactcaagt cccctgggtc aagtccttga ctccactctg 360
ggcatggcct aag 373

```

```

<210> 49
<211> 459
<212> ADN
<213> Homo sapiens

```

```

<400> 49
agtcaaaagg tgaatatata aacatgattt cagatagaat gcagaagctg tatttaatta 60
agttggcttt tgtcattaaa gaaatcagaa agtttggaca actcaggaat ggatttccat 120
cctagatggg atttactact atttctggaa ttaagtttac tgtctatatt caggattttg 180
tgcataagag acgtgaccac tgaggttggt tgtcattctt atctctactc caggggtcag 240
caaactttct ctgttaaggg tcagatagta gttttttggc ttttcagtct ctgtttggct 300
ctgctataaa caatacataa atgaatgagc atggctatgt tccaataaaa cttaattgac 360
actgaacttt gaatttcagt tacttttcaa atgtcacaaa tgttttccct ttgatttttt 420
tcagaaagcc attaaaaaaa tccaaagacc attcttagc 459

```

```

<210> 50
<211> 2575
<212> ADN
<213> Homo sapiens

```

```

<400> 50
tctggcattc tagagttctc tggcttgccc ttgttgcttt aaagctgggt ccaacttcgt 60
gggctttatg attgagcaac actgggtgatt gtgatttcta ataccaccct ccagattcat 120
gtcacagggt aaagtctgcc tgtggggaga gctttcacat tctcctggag ttgaaccaa 180
ataaggaaaa catcaacaaa cacagttacc aaaagcttcc aaggagggat ggaggagaaa 240
ttctgttcat tgcagatggg gagaaattaa atgacatcaa gtgcataatg gaactgggta 300
ggattcaaac tcaagtcccc tggttcaagt ccttgactcc actctgggca tggctaagtc 360

```

```

tggcctgtgt ctcctcttct cagctctagt ggctgtttag tagtgaagat actgatatgc 420
aggcagccat gttgttaaata agaagcaaat atttttatat ttctctctct tatgtgtata 480
atatttgagc taaaaacagg ttaaaacaaa agagagactt cagcttagta gcaggcaggg 540
aaggaagaag tgaatgcagc ttccttccgc ttacatttct aatcaagcta attatttcac 600
tagtgatttc tcatttgatt accgattttg cttatttatg cttagaactt tcaatgactt 660
ctcagtgcc aaggtaaa atgcaagcca tttctgtgac ctgcaatata atcttttgcc 720
actcctctc acttgctatg tcttagcaag actgagaggc tctcagttct tcagacctgc 780
tctatctcac ctctcatttt gtccatattc tccacctctt gtgcttgatg aactcctatt 840
cacctatcag gtcacaactt aaatatattt tctccaaat ctctggacta cataggcca 900
ggccccatc aaatgctccc atagcacatt gggcttctc atcaacatgc tcgttatact 960
tttctagaat tgcgtggctg tctccttctg tatctcccta atgcagagac cattactcat 1020
catttagca gatatttate catatcaggt actcagcaa catttgtgga atgaatctat 1080
atgtgaaact atacttctat ggcaagtttg tttggtcag aaggcttgca tttcatgtat 1140
attgggagac tggccaacat ttttatattc ttcctcctat gctcaaacca taaataatta 1200
tgtagcaaa cactgaactt tgtcntttga aacataaata gaaacatcct tagtaacagt 1260
actaacagta atcatttttc agaaataaga attcacacca gtgatacaga atggcctttg 1320
accattatct cagagaacta gcagtcgata gacatgctga agtgataatt ccaaaaaactt 1380
cagtttcatt catccttact taaatataag agaatcaact tctattttat gtctgtattc 1440
ccattttcta atatactact tgtcacagag gagatgccca accagtgttt ttcgtaaagt 1500
acctgaatag attaataata atgctagtga aatatattta gtgctgttct aagcacttta 1560
catgaattaa ctcatttaat cctcaccgca accataggaa gtagattcta ttatttatcc 1620
aattttacag atgaaaaaac tgaggctggt cacttgacat aaattaccag tgctggccag 1680
gacttttagg accatttaat ccaatgctct tattttacag ggggtgcagc gaagcctaag 1740
gaggtgatat gatttctctac acattcatgc acctcgtgc acatccatta aaacttttat 1800
taaatggaac ccactatcca tactccaatt agctaggtgc cttttttccc ccatttagta 1860
atttacaatt aattggagga ctttttatag ttatatattag aggaaaggag catgttttac 1920
tctatatcaa ctttcttttc tgcttaaaag cattttgatt ctttatttta ccagtcattt 1980
tagttctttt aataggttta atacacagat caaacaagac ttaaatacat ttttaacaat 2040
gtaacaattt aaaagaaatt gcctaaancc taggtcttct aacttggagc ccagtgcctt 2100
ttttattgct cgcgtgagtc aaaaggtgaa tatataaaca tgatttcaga tagaatgcag 2160
aagctgtatt taattaagtt ggcttttgte attaaagaaa tcagaaagt tggacaactc 2220
aggaatggat ttccatccta gatggtattt actactattt ctggaattaa gtttactgtc 2280
tatattcagg attttgtgca taagagacgt gaccactgag gttgtgtgtc attcctatct 2340
ctactccagg ggtcagcaaa ctttctctgt taagggtcag atagtagttt tttggctttt 2400
cagtctctgt ttggctctgc tataaacaat acataaatga atgagcatgg ctatgttcca 2460
ataaaactta attgacactg aactttgaat ttcatgtact tttcaaagt cacaatgtt 2520
ttccttttga tttttttcag aaagccatta aaaaaatcca aagaccattc ttagc 2575

```

<210> 51
 <211> 231
 <212> ADN
 <213> Homo sapiens

<400> 51
 tgattatttt cttttcccta tgttttagtta ctgtagtaga aggcctatatg aacctttggg 60
 aaaagttaga aggaaaatta acagtgtcag gcctcagtaa catatattaa gtttgtttct 120
 caaggagatc tgcctcgcct ttccttaaac atgttctgga tctgtaaaat aggggattct 180

tccatgaagg caaaagggtt gatttagaga gaaaagttta agcgtgggaa t 231

<210> 52
 <211> 344
 <212> ADN
 <213> Homo sapiens

<400> 52
 aatcccagaa ggaccttata ggtgagatac aaagtgttac ctgtgaggaa gtgtattcca 60
 ctccccacc aaaaaaatc tttggaatat atacaagcag aaatcgggag aatatgtgtg 120
 ggaatgccta ttaagggat aggataatgg tggagggaac ataaagattg aacaggccaa 180
 attattgata tattggccca ctaaacagga attatTTTT attgttgctg ctcatggagt 240
 tagaaaggtc tctagcagta tggttagttt gctgaaacat acatcaaaat ggtccaccat 300
 gagtgaattg gagatgttga acctctcttg gtttcatgca gagg 344

<210> 53
 <211> 402
 <212> ADN
 <213> Homo sapiens

<400> 53
 ttgttgattc tccccacc agcccctttg tttccagacc tataagtaga caagaatccc 60
 agaaggacct tataggtgag atacaaagtg ttacctgtga ggaagtgtat tccactcccc 120
 caccaaaaaa aatcttttga atatatacaa gcagaaatcg ggcgaatatg tgtgggaatg 180
 cctattaagg gtataggata atggtggaag gaacataaag attgaacagg ccaaattatt 240
 gatatattgg cccactaaac aggaattatt ttttattgtt gctgctcatg gagttagaaa 300
 ggtctctagc agtatggtta gtttgctgaa acatacatca aaatggtcca ccatgagtga 360
 attggagatg ttgaacctct cttggtttca tgcagaggaa gg 402

<210> 54
 <211> 228
 <212> ADN
 <213> Homo sapiens

<400> 54
 atttttctcc ctagtccatt tctttttgtc cagtattcct tttgttcata ggcagttcca 60
 ctgtgagtc tgtactcttt ctcatcatt catggctcac tgtgtgtcat gttctattct 120
 aggtgttggg gatatagcag tgaacgaaac agaagacccc tgacactgag gagctcatgt 180
 tctatgggga aggtgaagt gaaaaaaact caaaagaata aataaggg 228

<210> 55
 <211> 270
 <212> ADN
 <213> Homo sapiens

<400> 55

```

tgcctttaca atttcancgt tttttacccc caaatgggta atgtagggtgt tttttttttt 60
tttttgaatc attttattct ttcgtatttc tgtgtcagtc gctaggagaa agaccaggaa 120
aatctgttaa actctaaacc aaaaattccc attcactgca gagtgatgat tcgcaacact 180
gctgcaattc aggtagtggg gtgtcctgaa gtgacaacac atcaccactg cagtcctaata 240
aaattaagtc tttttgtttt gaaaaaaaaa 270

```

<210> 56

<211> 1130

<212> ADN

<213> Homo sapiens

<400> 56

```

accaattctc agtccttaga ccccatctct tgttctctct ctctctgatt ttcttcccta 60
gtccattttct ttttgtccag tattcctttt gttcataggc agttccactg tgagtcctgt 120
actctttctc attcattcat ggctcactgt gtgtcatgtt ctattctagg tgttggggat 180
atagcagtgga acgaaacaga agacccctga cactgaggag ctcatgttct atggggaagg 240
tgaagtggaa aaaaactcaa aagaataaat aagggaattt caagagggtt gaaggaaatc 300
taactaggta atgtgattgc gggacctagg gatagggaag ggcatagact gcttttagata 360
ggatgggaaa tctcttaggg gagattaaaa ttgaactaag acctgaatta catggaaaaa 420
tttagtcaaa ggaagatcat tccagagaga gagcatggca gtgcaaagcc ttgcagaaag 480
aaccagtgtg gaaagtccaa ggatgaaaga aagccagtgt agctagagcc tgaagaatga 540
ggatggggagt ggaagatgtg aaggtaagtt tgcagggggc agatcatatc aggcctggaa 600
gatgatatag ctgttttaaag tttattctaa atgcaatgag aagccagctc ctagcaagat 660
aataaattgc ttttccagaa cgttaactgt ccattattta gttcctttgc attctccata 720
ggctggtctt ataaagggtt tgtgacagta gggaatccat taaagatcta ttacataagg 780
gaatgtggga catgcaaaag cagaagacaa atgtagaaac cctctgcttc acttggcatt 840
tctgtccttc atctttctga ttttttaaag ctactgtgtt cctacataca tatgctatct 900
gttccttcat tctttttgat aaattcatca ttgacctta catttttgc gttttttacc 960
cccaaagggt taatgtaggt gttttttttt tttttggaat cattttatc tttcgtattt 1020
ctgtgtcagt cgctaggaga aagaccagga aaatctgtta aactctaaac caaaaattcc 1080
cattcactgc agagtgatga ttcgcaacac tgcgtcaatt caggtagtgg 1130

```

<210> 57

<211> 463

<212> ADN

<213> Homo sapiens

<400> 57

```

cctaattgagt ggctctgtat tcccatttcc tcatgtggac taatgcgggtg tgtggtgttt 60
ttcttttccct ttccagaagc agtggacagg ggagaatccg ctccaacgtt gtccacctcc 120
ccttcacctt cctcccttcc acccacttcc ccttcacctt ctctgggcag acgaaggcct 180
gnaaatagga acgtttctta gaaagaagaa aactagtga atctactttg tgtctctagt 240
ttgggccatt gtcgtcatgc agatctggtt gaacctgtgg attgtgcagt ttgctgccag 300
tgccgattgc aggtcgttat ctggttnaag tacgatctcg tcttgtcttt ggaagattga 360

```

tttttttttt tttttttttt tnggatttta atntttgagn gacattttac cgtnttttgea 420
 aagtaaaaca ttttgatggg ttggnatttt aaatagttna ctt 463

<210> 58

<211> 289

<212> ADN

<213> Homo sapiens

<400> 58

gactacagtg gagagcagct tccaaaaatc agattacgtt ttaatcaata gtgaaaaacc 60
 ttgacatagg aaatgaggag taatcaaata tagaaggtag tcagaaggca aaatccacag 120
 cagccactaa ggacttaata aatttgggca caaaatatgg gctgccataa ggaagagagc 180
 ccaaataatc aggaataat tagggtaaaa gatgtttcta caatgcacct gtcaggggaa 240
 aactctccct tgctctttct agtcttaggg nggaaggaca gttttaggg 289

<210> 59

<211> 1542

<212> ADN

<213> Homo sapiens

<400> 59

tgggccattg tgcgtcatgca gatctgggtg aaactgtgga ttgtgcagtt gctgccagtg 60
 ccgattgcag gtcgttatct ggtaaagtag gatctcgtct tgcctttgaa gattgatttt 120
 tttctttttt tacttttggg ttttaatat tgagatacat ttaccgtctt gcaaagtaaa 180
 acaatcttat gtttgcattt aaatagtaaa caaacaccat ttggttaatt tattactttg 240
 atacatcttt aatatttcac tttttttttt ttttttgaga cggagtcttg cccaggctgg 300
 aagtgcagtg gcaccatttc agctcactgc aagctccacc tcccagggtc atgccatttt 360
 cctgcctcag cctcccaggt agctgggact atagggtgcc accaccacac ccggctaatt 420
 ttttttttat ttttagtaga gacagcgttt caccatgtta gccaggatag tctcgatctc 480
 ctgacctcgt gatctgccc tctcggtctc ccaaagtgtt gggattatag gtgtgagcca 540
 ccgcgcccgg ccagtatttc acttttaagc tcaaaaaatg tgttttagctt cttacaattt 600
 catgtagtat ttgagattaa tagctccata gagagatttg attgttttagc catccatgat 660
 tctgtgaagg agaattgtta caatatctgt gttttggaat ttgaagatcg aatgattttt 720
 tgtttttttt tgagacggag tctcgtcttg tcacccaggc tggattgcag tggcgcgac 780
 tcagctcact gcagcctttg cccctgggt tcaagcaatt ctcccacctc agcctcctga 840
 gttgctggga ctacaggcat gcgccaccac ccctggctaa ttttttgat ttttagtaga 900
 gacggagttt caccatgttg gccaggctgg tctcgaatcc ttaccttaag tcatccgctc 960
 acctccgctt cccaaagtgc tggcattaca ggagttagcc actgcgccta gctgagattg 1020
 aatgattttt gagcgtact cttagggaag cacgttttac cttgcctttt taattcaaag 1080
 gaaatgtagg ctttgagaaa tacatgtatg caaagtcttg tgatttggca gaaaaattaa 1140
 attacatgta tatctatatt atatgtatta catcaaaatg gaatcatagt tttccataat 1200
 ccattaaatc aggtctttgt atcttaatgc tgtggtaatt taaggccac ctctattgtt 1260
 ttgattctct ctaaaactgt ctttctccta agactagaaa gagcaaggga gagttttctc 1320
 tgacaggtag attgtagaaa catcttttac ctaatatatt cctgattatt tgggctctct 1380
 tcttatggca gccatatttt gtgccaaatt tattaagtcc ttagtggttg ctgtggattt 1440
 tgccctctga gtaccttcta catttgatta ctctcattt cctatgtcaa ggtttttcac 1500

tattgattaa aacgtaatct gatttttggga agctgctctc ca

1542

<210> 60

<211> 1417

<212> ADN

<213> Homo sapiens

<400> 60

```
gacacggctg gctgcttttc tcagcgccga agccgcgcca tgctcgtcct cagaagcgcc 60
ctgactcggg cgtggcctc acggacgctg gcgcctcaga tgtgctcatc ttttgctacg 120
ggaccagac aatacgatgg aatattctat gaatttcgtt cttattacct taagccctca 180
aagatgaatg agttcctgga aaattttgag aaaaacgctc atcttcggac agctcactct 240
gaattggttg gatactggag tgtagaatgt ggaggcagaa tgaatacagt gtttcatatt 300
tggaagtatg ataattttgc tcatcgaact gaagttcgga aagccttggc caaagataag 360
gaatggcaag aacaattcct cattccaaat ttggctctca ttgataaaca agagagttag 420
attacttate tgggtaccatg gtgcaaatta gaaaaacctc caaagaagg agtctatgaa 480
ctggccactt ttcagatgaa acctgggtgg ccagctctgt ggggtgatgc atttaaaagg 540
gcagttcatg ctcatgtcaa tctaggctac acaaaactag ttggagtgtt ccacacagag 600
tacggagcac tcaacagagt tcatgttctt tgggtggaatg agagtgcaga tagtcgtgca 660
gctgggagac ataagtccca tgaggatccc agagtgtgtg cagctgttcg ggaaagtgtc 720
aactacctag tatctcagca gaatatgctt ctgattccta catcgttttc accactgaaa 780
tagttttcta ctgaaatata aaacatttca ttaactgcta taggatctgt ctgctaattg 840
tgcttaaatt ctccaagag gttctcactt ttatttgaag gaggtggtaa gttaatttgc 900
tatgtttctt gcattatgaa ggctacatct gtgctttgta agtaccactt caaaaaatag 960
ttctgtttac tttctgcatg gtatttcagt gtctgtcata cattaaaaat acttgtcact 1020
gttttaagat ctgactctt catttgtttc agaatagctc ttctactgta ttctgacaac 1080
tctttgcttt atagcatttt gttgtattca aatgataatg gtagcatttc catgcttgtg 1140
acagcatttt taagtattta atatatatta tcaacctttc catcatgtct gttttcctgg 1200
tttttttttg ttgttttttg accagtaaaa tttattttgt aataccaaat aggatttaag 1260
aaaattaacg tatttcttta ctatggaaaa ccacattgtc atttgtgaca tcatctatat 1320
taaataatgt tttcacatta gttatttgtc acttacttgg aaaatgatgc tgtaggttcc 1380
tggtattaaa aatctagaaa agacttggtg gtttatg 1417
```

<210> 61

<211> 696

<212> ADN

<213> Homo sapiens

<400> 61

```
ctctaagcga tctgcagttt atctgggcta ttatgcaatg tgagatatta tttatgagag 60
gcaggcagtg gcagcttaaa cagacatacc tcttcagcat tcttttccat gcagagttca 120
aagaggccta agacatctca catttatgtg aaatagttga ggcacaagta tgtgggttaac 180
tacacataaa ttaccaataa aattaaaagt taactaaaat actacagaag acatttaata 240
ttttgcttca gactgttata gtaccttagg gttatgaaca tcaattacct ttgagcctta 300
ctatatttaa aggggaagggg aagctgacat ttttatcttt tttaatgggt acactctacg 360
ttaccttaaa ggcttccccg gtttaagcaa aaatagaatt atgtgggtta aatcatatac 420
```



```

cttattcttc atacttagct ccaagtagat tatgactttt cctaaaactt aacatgacag 480
agatttgcct tccagggggt ctaatttgaa gaggaagta atttaaagt ataggtttat 540
ataatagcat tgtatttttc tctattttat tcattcttcc attctcttcg atgaactcca 600
gcacctgcgc gtccatcttc attcttagcc aatgacctt ccttttactt aacagagaaa 660
ataacaatca gaagagtctt cataagttcc caccoc 696

```

<210> 62

<211> 2702

<212> ADN

<213> Homo sapiens

<400> 62

```

ctcagaaagg acacggctgg ctgcttttct cagcgccgaa gccgcgccat gctcgctctc 60
agaagcgccc tgactcgggc gctggcctca cggacgctgg cgccctcagat gtgctcatct 120
tttgctacgg gacccagaca atacgatgga atattctatg aatttcgttc ttattacctt 180
aagccctcaa agatgaatga gttcctggaa aattttgaga aaaacgctca tcttcggaca 240
gtcactctg aattgggttg atactggagt gtagaatttg gaggcagaat gaatacagt 300
tttcatattt ggaagtatga taattttgct catcgaactg aagttcggaa agccttggcc 360
aaagataagg aatggcaaga acaattcctc attccaaatt tggctctcat tgataaacia 420
gagagtgaga ttacttatct ggtaccatgg tgcaaattag aaaaacctcc aaaagaagga 480
gtctatgaac tggccacttt tcagatgaaa cctgggtgggc cagctctgtg ggggtgatga 540
tttaaaaggg cagttcatgc tcagtcaat ctaggctaca caaaactagt tggngtttcc 600
cacacagaat atggagaact caacagagtt catgttcttt ggtggaatga gagtgcagat 660
agtcgtgcag ctgggagaca taagtcccat gaggatccca gagttgtggc ggctgttcgg 720
gaaagtgtca actacctagt ttctcagcag aatatgcttc tgattcctgc atcattttca 780
ccattgaaat agttttctac tgaaatacaa aacatttcat taactgctct aagatgtgtc 840
tgctaattgg gcttaaatcc tcccaagagg ttctcgcttc tatttgaagg aggtggtaag 900
ttaattagtt atgtttcttg cattatgaag gctacatctg tgctttgtaa gtaccacttc 960
aaaaaatagt tctgtttact ttctgcatgg tatttcagtg tctgtcatac attaaaaata 1020
cttgtcactg ttttaagatc ttgactcttc atttgtttca gaatagctct tctactgtat 1080
tctgacaact ctttgcttta tagcattttg ttgtattcaa atgataatgg tagcatttcc 1140
atgcttgtga cagcattttt aagttattaa tatattttat caacctttcc atcatgtctg 1200
ttttctgggt ttttttgggt tgttttttga ccagtaaaat ttatttttga ataccaata 1260
ggatttaaga aaattaacgt atttctttac tatggaaaac cacattgtca tttgtgacat 1320
catctatatt aaatatgggt ttacatttag ttatttgcct cttacttggg aaatgatgct 1380
gttaggtcct ggtattaaaa atctagaaaa gacttggttg ttatgtgct gaaatgtntt 1440
tntttataat taattttact aatttctact tagtttggat cactaacaga gatcttggga 1500
catttatttg ttttaagaa atatttatgg ttatggaaac gcttgcccta ataaaaatcc 1560
tgcataattc ttgttttttt aaattcacat ttatactta tatgatctct aaagctcttg 1620
ctatgttgct ataagacagt aatatagtga taatttacca actttattga aaatgttggt 1680
acatcaataa aatagcatgc tgggaacctg agaaggaagg tttcttttagt actgccacag 1740
atgtcctgga gttttccttg aaacatcttt catttccatt ttacttgatg cctctacctt 1800
ctaaaactag atttggtact tccgatagtg tatattttaca gctatattat atttacattt 1860
ttacagacta tcttctgata atctgtattt acagctatgt tattttatat ttttaaaatc 1920
ttacagctat aaaatagtaa attgtgaatt agctgtccaa gggcaatttg gaaaacaaac 1980
tgtacagttt cctctcattt atcaggctct aagcgatctg cagtttatct gggctattat 2040
gcaatgtgag atattattta tgagaggcag gcagtggcag cttaaacaga catacctcct 2100

```

```

cagcattctt ttccatgcag agttcaaaga ggcctaagac atctcacatt tatgtgaaat 2160
agttgaggca caagtatgtg gtttaactaca cataaattac caataaaatt aaaagttaac 2220
taaaatacta cagaagacat ttaatatattt gcttcagact gttatagtac cttaggggta 2280
tgaacatcaa ttacctttga gccttactat atttaaaggg aaggggaagc tgacattttt 2340
atctttttta atgggtacac tctacgttac cttaaaggct tccccgggtt aagcaaaaat 2400
agaattatgt gggtttaaac atatacctta ttcttcatac ttagctccaa gtagattatg 2460
acttttccta aaacttaaca tgacagagat ttgccttcca ggggtctaa tttgaagagg 2520
aaagtaattt aaatgtatag gtttatataa tagcattgta ttttctcta ttttattcat 2580
tctttcattc tcttcgatga ctcccagcac ctgcgcgtcc atcttcatto ttagccaatg 2640
acctttcctt ttacttaaca gagaaaataa caatcagaag agtcttcata agttcccacc 2700
cc                                     2702

```

```

<210> 63
<211> 335
<212> ADN
<213> Homo sapiens

```

```

<400> 63
taccttacca ttgtgcctgt tatgttcaaa aatacatttt gaaaagttct taaaataatt 60
gctctaaatt gaaatgtcaa aacctgaata ttctacatgt gtcttgggtg atgtccttta 120
agctaatect atantacact aagggtacatt taattgtgga aatgtcattt cacaaggcta 180
tgtaaggcat cctcgcaggc ctgcatttta atgtctagtt ttcaatttta tggcacctaa 240
aatatccena tgggtggagga tattagggat ggggntatta tcccnggat ntctacngg 300
cccggttggg gccccgggga tcttaggggt tcca                                     335

```

```

<210> 64
<211> 447
<212> ADN
<213> Homo sapiens

```

```

<400> 64
tccttactgg attgagcctt gggtacttgc caccagagg gtgattgagt tgcagtattc 60
tgttgcacct ttgatgtcaa cttatttttg gcttaaattt tttaaataac ataagggtta 120
agaaaggatc acatctctat cttttcctca ttgacctccc tagttctaga caaattctat 180
acttttttcc tttctttttc tctgtggtcg attatgtcat tatccctcc cctcaaaaac 240
ctaccttctc ccttgggtctc cggaataact gcactctagt gattttcttc ctactacgga 300
aatgatcttc attacctctc tgggtctggg atttccttgg tctttagggc cctttatgga 360
aaacacaacc tttggcctan tttctctnca atttgccctt ttaagaatg gtcttggtac 420
tccttttata attaactctg gtgnggc                                     447

```

```

<210> 65
<211> 2324
<212> ADN
<213> Homo sapiens

```

<400> 65

```

actacaagcc agtgaagtgg gtgataagcc tgactccact atctcagcca ggtccttctt 60
ctaataattat tggccagtct gtggaagaag ctatcagagg ggtgtttgat gcttccctca 120
aaatggctgg cttctatgga ttgtatacct ggctgactca tactatgttt ggcatcaata 180
ttgtcttcat accatcagca ttagcagcaa tccttggagc agtgccattc ctggggacat 240
actgggcagc agtaçctgca gttcttgacc tgtggctgac acaagggtta ggatgcaagg 300
ccattttact gttgattttt catctcttgc caacatactt tgtagatact gcaatctact 360
ctgacatatac aggaggtggc catccttacc tgacaggctt ggcagtggcc ggtggagcat 420
actacctagg cctggaagga gcaatcatcg gtcctattct tctctgcata cttgtggttg 480
cttccaatat ctatagtgcc atgctagtga gtcccacgaa ttcagtcccc acgccaacc 540
agaccccatg gcctgctcag cctcagcgga ctttccgtga catttctgaa gatctgaaat 600
cttcagtagg ttgatgtggt tctctctgca gtgatttttc taggaagtcc aaatttgaca 660
gcgagtccag ctcagctgtg gccctctgcc cttccagctg tgccatagca gcaaaaccca 720
ggaaagaagc agaagcctcc tggccttaca tacagaatgc ctggacaaga gagaacttgc 780
tgcgggctgc tttgtatttt aaaacacagc ttgagagtcc agagtgggtg gtttgcacac 840
ttaactgttg ttaagatggc ttgaaaagtt tcattttata cactgggtacc ctggcttgaa 900
atttttccac tttggttatc tatgttacta tattatataat ttataaagtt attttaagaa 960
ctctaaacta cctgctgtta aaagaataga tgggtgtaatt ttttctggt ttaagaaatg 1020
tattgttaaa cttttctaag acagtcactt ttcaaggaag agggctttca cttttgagt 1080
tgtagttgag tgagcaggaa aaatgaatct tctacccttc tcccacaatg tattatacgc 1140
tctttaagaa ataataaatc ataagtataa ggggtgggtg gcttatttga tgttcagttt 1200
tatattataa cctggggaga tacaaagact gaagctcttc ctcctcttcc ttttctcttc 1260
acgcttcttt acactattgc caaattataa aacttggctg accacgttga agtgaaatac 1320
ttattaagct gctatgaatg gtaacagtat gataaaatc atgctgttat taggttttct 1380
ccttccaggt ggttgaagtg aaaaatctca ggtgtagcaa tactctgatt tgttaaagt 1440
atccgtttca ttatttgga ctctgccaat aagttagttt tcagcagaat tttgtattta 1500
tgtagtattt tcccatctct tagcacagt ccttgcacat gactctaaat gaatgcttgt 1560
tgaattgaat tgcaacttta aaaatatctc agatatcaca ataaccagca tagtgcttta 1620
cattttacaa agtccttttg cacatactat ttcacaagta catcacatca gctttcatat 1680
agctttatat tattttacat cagttttaca ctgatggaga agctaagggt cagaaaagt 1740
accaactcaa agtcgtaaag tgtagtagga tttggcctca gatctttgac tccgttgtgt 1800
gcagattctg ctgtattagt gttccttcta aatattgtat gcagttttcc cagtgttacg 1860
agattgggtg tgctatccat ttgttgtgaa tgaagacttg tcctatagta catgacaggg 1920
ctgagagtgg actgtagtcc ttagctgaac cctttctact gtttaaaaca atcatcaggc 1980
ctatcctctg ctcccagggg aatgtgggag agaaaagggg tcttggccag gctgtagggt 2040
gtggctcaca tctgtaatcc cagcactttg gaaggctgag gtgggtggat tgcttgagcc 2100
cagggatttg agatcagcct aggcaacatg gcaaaacccc atctctacaa aaaaaaaaaa 2160
aaaaaaaaaa aaattagctg ngnacgggtg tgtggctata gtctgagcta ctcaggagggc 2220
tgatgtatga ggatcacctg agtccaggag gtcgaggctg cagttagcca agatggcgcc 2280
actgcactcc agcctgggtg agagagttag acactgtccc aagg 2324

```

<210> 66

<211> 447

<212> ADN

<213> Homo sapiens

<400> 66

```

tgttggacct ggtactctcc accacataat tatgtcctgg tacttagggg agcctggcta 60
attcttttca caacatttga aaagccgtac gtttaacttc atggatgtgg agaaacagca 120
taatgacctt gggtcacaaat gaaagtgtca gccttctttc tcgtggctgg aataagaaat 180
gcaagaactg gaagattaac agagagctga taccgggact ggtaatcca caactcacta 240
ttttgggagg aaggggaaga caggggaaaga gaaggaaaaa gggcaaatgg ggaaacattt 300
cccagagct gtgntcctgg gtcttcacct gcaaaaaaga tttgcagcca cctaaggaga 360
aagntgagct tgccntcag ttccctcantg ttcacctacc ttttttcaca acatttgta 420
cccaggttga gggtttttcc agatttt 447

```

<210> 67

<211> 564

<212> ADN

<213> Homo sapiens

<400> 67

```

attgggcctc ggnncaaaa aacgggcacg angcggactg ggtaancca naactcacta 60
ntttgggagg gaaggggaag acaggggaaa gaganggaaa aaggcaaang ggaaacatnc 120
cccagagctg tgcncctggt ctccactgca aaaaagaatt gcagcccacc ctaagagaaa 180
gctgagctgc cccctcagc cctcactgtc accctacctc tatcacaaca attgtacca 240
agatgaagac ttctcccaga cttcagaaaa taaagtcaaa ccctagattt tgttttaaaa 300
taggaaactc agaatcaact tgcctccatc ctctgggaaa actgctccca cacaggcctt 360
ggagtgtgtt gtagcactgt ggaggaatgc agaaaggatg aaagagatct tgattctcct 420
agtggttctc ttcactaccg taggcattcc tcagccattg actcctcctt ctttctcttg 480
acattcactt tcttggccag tcttacctgc ttatgagtct actttccaat aaatttactc 540
atagtcenat taacttcna aant 564

```

<210> 68

<211> 604

<212> ADN

<213> Homo sapiens

<400> 68

```

cacataatta tgtcctggta cttagggtag cctggctaatt tcntttcaca acatttgaaa 60
agccgtaggc attaaacttca tggatgtgga gaaacagcat aatgaccttg gttcacaatg 120
aaagtgttag ccttctttct cgtggctgga ataagaaatg caagaactgg aagattaaca 180
aagagctgat acccgactg gtaatccac aactcactat tttgggagga aggggaagac 240
agggaaaagag aaggaaaaag gcaaatggga aacattcccc agagctgtgc tcctgggtctc 300
acctgcaaaa aagaattgca gccacctaag agaaagctga gctgcccctc agtcctcact 360
gtcacctacc tctatcaca caattgtacc aagatgaagc cttccccaga cttcagaaaa 420
taaagtcaaa ccctagattt tgttttaaaa taggaaactc agaatcaact tgctccatc 480
ctctgggaaa actgctccca cacaggcctt ggagtgtgtt gtagcactgt ggaggaatgc 540
agaaaggatg aaagagatct tgattctcct agtggttctc ttcactaccg taggcattcc 600
tcag 604

```

<210> 69

<211> 482
 <212> ADN
 <213> Homo sapiens

<400> 69
 aaaaagctct gccagctgag cgnacttaaa ttgaggaaga gttcccttct ctattaactc 60
 attttgggca gtaacaacat accctgcctg atattttcct tctgagactt tggcatagga 120
 cccacgaaca aaaagtatca actcaggatt atccaatgga gcatctcata aatcaacaca 180
 aggggctact agtttgata ctacacctac acaattgtga tctttaccac agtcagtcag 240
 aggtaaatgga gtagcagggt taagtagggt ttaggatggg agtaggcatt ttaggatggg 300
 aggtcagaag ggaggnccagg gaggaataatt tncattaagg ttcataagggt gttagttttc 360
 tttggctgga aaantgctgg agcttgggtt gggatttttag gtaggacttt tccacagcct 420
 gtgggaattt tggcaantta agtttcnttc ccttaaaacc cgttccgggtg gaaanttggt 480
 gg 482

<210> 70
 <211> 402
 <212> ADN
 <213> Homo sapiens

<400> 70
 ttccagntat ggatatgtta cagacatgac tgaattgggtg gccttagctg aacattttga 60
 gaggactctg gagcaagaaa aaacccaaaa gactaccaag cttatgtctc accaattaca 120
 acagttacag gggctgagac caaaggaacc ttctttttta ttttaaatca caaccaagag 180
 atagatccta ggaacaagaa attcttttcc cggaggatgt ctgccttcc tggcaaataa 240
 ggcaggagaca ctgggaaaag agttgtctgc ttctctgtta ggtctacca taaggcctcc 300
 tcccattagg gccacactgt ttcaccacta aagggaaggc ctaaggacag atcctgtgtc 360
 tttaaaaaaa aaaaaaattg gaggcggcgg caggccacca ng 402

<210> 71
 <211> 758
 <212> ADN
 <213> Homo sapiens

<400> 71
 agtaacaaca taccctgcct gatattttcc ttctgagact ttggcatagg acccaggaac 60
 aaaaagtatc aactcaggat tatccaatgg agcatctcat aaatcaacac aaggggctac 120
 tagtttggat actacacctc cacaattgtg atctttacca cagtcagtca gaggtaattg 180
 agtagcaggg ttaagttagt tttagatgga gtagcatttt agatggagggt cagaaggagg 240
 caggagaaaa tttcataaga tcataagatg ttagttttct tgctgaaaaa tgctgagctt 300
 ggttggaatt tagtagactt tccacagcat gtggaatttg caaattaagt tcattcccta 360
 aaaccagctc agatgaaact tgtagcaact tgggtggctgc tgctgcttct attttttttt 420
 ttttttttaa gacacaggat ctgtcttagg ctctctttag tggtgaaaca gtgtggccta 480
 atgggaggag gcttattggg agactaacag agaagcagac aactcttttc cagtgtcctg 540
 cttatttgca aggaaggcag acatctcggg gaaaagaatt tcttggtcta ggatctatct 600
 cttggttggt atttaaaata aaaaaagaagg ttcttttggt ctcagccctt gtaactgttg 660

taattggtga gacataagct tggtagtctt ttgggttttt tcttgctcca gagtcctctc 720
aaaatgttca gctaaggcca ccaattcagt catgtctg 758

<210> 72

<211> 673

<212> ADN

<213> Homo sapiens

<400> 72

caggttgagc ccttgaaaaa cacaaaaagg acaagttcct gtgttgaaac acccccatcc 60
cctcacgttc tcacccccag gcccagaagt aggttgcaac tgcctttgga agattttgcc 120
ccttagccat cccacccac ttgtaccagc taagaatgct ggagactctg ccaccatgct 180
ctgcgtgccc ctgaacctct gtgcagcccg gaaggctgat gtacaggtgt acctcaatcc 240
acattacagc catgctccta atgtacatgg acatttttgt aactcagctc atattctgac 300
tgtatttgag aagctggctg ttttaagggaa cccagaagtg aattcttttg taaagtaaag 360
cacccttttg taatgcaatt aattatccct taatgtatct gttttgtaag tctgcatttt 420
tgtatatcgg atttacctta agcttctcta gtgaggcatt ctgagcagtg gtgatcacat 480
gccagatccg gcctgccna tccacaaaag tagatganca atgcacgctc ctcaaacatc 540
tttggaggaa ctacctgggc canancactg gccaggggtg cagcaagcag cagcaggggc 600
tgacagcagg gtttactgnc attcaacatt gcttganatt gcctctaatag ttctgnatta 660
aangaaaaac cat 673

<210> 73

<211> 554

<212> ADN

<213> Homo sapiens

<400> 73

aaagggggaa aggaaactac ttttaagagcc ccctttttcn tccnaactc caaatttttg 60
aaaggggaaa aaacnggaaa accccacccc agttccctcc agaaaaatgc cttttcctag 120
gncattaaat acctaccttt acctatccta anactgnnta agntaaatcc ctttccttta 180
ccccacccta aactanccca ttctaانccc ggantccctc actctttttt tttagttact 240
aatcattttta tgaataaat gtattttataa gtattttcct taaggtttgt gaagagtatt 300
tgcantgtgt ctttcatttt natgtgtttg cnannnetcc gctccaggaa gaacggaaat 360
gctgtcttgt gagcatgaag tgaacgggct gttttgctcc agccactttt cttgtacaac 420
cacatggatg gattagatgt cctcaggtct tttccatctt cagtttctat gactgtggaa 480
taaattgttca gatagaaact tcaaaaaaaaa aaaaaaaaaa aaaaaaaaaa gtacctcggc 540
cgcgaccacg ctaa 554

<210> 74

<211> 1794

<212> ADN

<213> Homo sapiens

<400> 74

```

gttttcaccc caggcccaga agtaggttgc aactgccttt ggaagatttt gccccttagc 60
catccccacc cacttgtacc agctaagaat gctggagact ctgccaccat gctctgcgtg 120
cccctgaacc tctgtgcagc ccggaaggct gatgtacagg tgtacctcaa tccacattac 180
agccatgctc ctaatgtaca tggacatttt tgtaactcag ctcatattct gactgtattt 240
gagaagctgg ctgtttaagg gaaccagaa gtgaattctt ttgtaaagta aagcaccctt 300
ttgtaatgca attaatatc ccttaatgta tctgttttgt aagtctgcat ttttgtatat 360
cgggtttacc ttaagcttct ctagtgaggc attctgagca gtggtgatca catgccagat 420
cgccctgcct atccacaaag tagatgacca atgcacgctc ctcaaacatc tttggaggaa 480
ctacctggcc aaaacactgg ccaggatgca gcaagcagca gcaggggctg acagcaggct 540
tactgccatc aacattgctt gaaatgcctc tatgttctga ataaagaaaa accataattg 600
cttgtggtga aacgaagcag tcttcatgtt aagtagcaat ggttattttt attggtagta 660
actgaacagt gttttgcaat ttgtgaaaca gtgtattgtg ttttgtaaaa tgatgtcatg 720
aaatggtggg tccttggaac cctcctttcc gttcagctct gcctctgttc tttcaactcc 780
tttgaggctc aaaaaaaaca caaagatcag aagccttcag atagagggtg gtattctggt 840
aaagaagaaa gagataaggg acgctacctt gcttttctgg cacaggaagc acatgataaa 900
gcatgctcag atgagctgga acagatatag ctacctggtt cgtgtaaata agaataatca 960
aggccccaga gtgtgtatgc ttccagggtg aggagaaagg ggaatctccc aaaatttaaa 1020
aacaattgg aagaataacc aggacagcca agtgaagcag ccacagggac ccaagcagtc 1080
gaggtcttta atgtgcctgg agatgactct ctgctattca tgaatcttgc tattgcacaa 1140
accctatcaa gagctgctgc ttcccttcca gccagaaaag tggtaagcgg agcaagtgcc 1200
aagcagaaca gaccttatca tctgggtaac agacttctca gtgttggtgc tgtgtctggt 1260
agagccttag agcaagttaa gcacttcctt ggtgtgggta aagaataaag gggaaagaaa 1320
ctactttaga gcctcttttt ctcccaactc atatttttga taggaaaaac agaaaacca 1380
tccagttctt cagaaattgc tttctaggca ttaatactac tttactatct atactgttta 1440
gttattcctt tctttacca cctaaactat ccactaate caggattccc tcaactcttt 1500
tttttagtta ctaatcattt tatgaaaata atgtatttat aagtattttc ttaaggtttg 1560
tgaagagtat ttgcattgtg tcttcatttt aatgtgtttg caatcgctcc gctccaggaa 1620
gaacggaaat gctgtcttgt gagcatgaag tgaacgggct gttttgcaat ccagccactt 1680
ttctgtgaca accacatgga tggattagat gtccctcagg cttttccatc ttcagtttct 1740
atgactgtgg aataaatgtt cagatagaaa cttcaaaaaa aaaaaaaaaa aaaa 1794

```

<210> 75

<211> 507

<212> ADN

<213> Homo sapiens

<400> 75

```

aagaaatgat tgaaatgttt caaggggaca tggagctgaa aggtggaatg agagacttcc 60
gtttgggcta taagcttgtg gttttatttg accttttaaa ctgagcactt acactacttt 120
atttctaaat gttctatgaa ggaataagag ttgtagaac aaaagagaag ggctaaaaga 180
gagaccagtt agaaagaggt ggagagtcag aactggagtg caaaagcaca aagattccag 240
ggcatacaag gatgataaat cacttcaca aatgagcgca aagggatatt gtgtacaatg 300
cttccatgaa ggaaagaggg gcaactggatt aggtaccac ctgggtttga attgaacaac 360
tactgttttt cccaagcttt tgtgtttgac agtttttaac cttttggcat gttccatttt 420
cccggggngn ttgaggacn tttaggagct tgggatgcct taagtcccgt taacnttaag 480
ttttctttcc ccccaagggn gggtttna 507

```

<210> 76
 <211> 415
 <212> ADN
 <213> Homo sapiens

<400> 76
 ttttttttaa tatttcgatg acatgttgaa tgcataataag aggttagagt gaacaaggaa 60
 gttaagggct ttcttttttc aaaatcatat ttgcatttaa cttgaaattc tttccagtca 120
 tacgattttt gcataaatat ttatgggatc ttgatggcta attagattat aagtctttta 180
 tccaggaatt tgtctaattt tatgctctcc atcaaacctg agctctgtgc tgatcttcag 240
 tccatgttaa acaaaattat tttaacaccc ttttaaggctg catttcaa at tctctcattt 300
 gttactcata ctatgtcatg ggaatatata gtctttattt ttgggatgaa agtttgaaga 360
 cccaaatccc ccaaaaattg gatttcccnt gtggaaagct ctcatTTTgg atcca 415

<210> 77
 <211> 1318
 <212> ADN
 <213> Homo sapiens

<400> 77
 ggagctgaaa ggtggaatga gagacttccg tttgggctat aagccttgtg gttttatttg 60
 accttttaaa ctccagcactt acactacttt atttctaaat gttctatgaa ggaataagag 120
 tttgtagaac aaaagagaag ggctaaaaga gagaccagt agaaagagggt ggagagtcag 180
 aactggagtg caaaagcaca aagattccag gcatacaagg atgataaatc acttccacaa 240
 atgagcgcaa agggatattg tgtacaatgc ttccatgaag aaagagggca ctgataggta 300
 cccacctggt tgaaatgaac aactactgta ttccaagctt tgtgttgaca gttttaacct 360
 ttgcatgtca ttttccagggt agtgagacct ttagagctgg atgcttaagt ctgtaactta 420
 agtttcttcc acaaagagga tttattttgc tttttgtgac atgatgggtgc tagaacaaaa 480
 tgtactaaga aaaatttgat ggcagtgata gcattatctt gagggataaa gaagtggagt 540
 gagaatgcta agtgtagatg aaaaggcctc ttcaaaaccc atttggattt gtaaagtgc 600
 tgcttgaga cagcttgagc ctctagaaat gagaagaaca atgggagaat ggttttattc 660
 ccactgaaat aattgcaagt tgaccttctt ccaactagaa aagtgggggt gtgtctgagg 720
 acttggttg agaataattt ttcttagtt gcttcatgat actgaacagg tcaactagcc 780
 tttctctcc ttactttctt tcacagttaa atgagatggt aatctccact tctaagggtt 840
 gttgtggaga ccacatgaag acattgatgt aactctgcct ggcagatgat gtaaagtgcc 900
 catttgtgcc ttttgccttc tttgcttttg aggcatttgg cacttttatt attccttgat 960
 caaaatgaga gctttcacat gggaaatcaa tttttggggg atttgggtct tcaaactttc 1020
 atccaaaaat aaagactata tattccgtga catagtatga gtaacaaatg agagaatttg 1080
 aaatgcagcc ttaaagggtg ttaaaataat tttgtttaac atggactgaa gatcagcaca 1140
 gagctcaggt ttgatggaga gcataaaatt agacaaattc ctggataaaa gacttataat 1200
 ctaattagcc atcaagatcc cataaatatt tatgcaaaaa tcgtatgact ggaaagaatt 1260
 tcaagttaaa tgcaaatatg attttgaaaa aagaaagccc ttaacttctt tgttcact 1318

<210> 78
 <211> 530

<212> ADN

<213> Homo sapiens

<400> 78

```

tagtctctac aatattgtca gactgatctc tgcaaaactc aaaagtcaaa ttactcttct 60
actttaaaact tccagtagct cccggattca ttttaagggtga gaatctaaat cctttggcat 120
agcacttaag accatttcca agggagtggc tctaccatgg agccctagag actctacatg 180
tgtctacatg ggatgcacag gcaaagcttg gactgccata tgacctaaat ctgggtggaa 240
tgccctgtga tcctcctagg aacatgggca agggaggaaa tcttcagaga aggcttcgcg 300
agggtgtttt ctcatctgcc tctatcttgt ggggatgctg ccatgaggcg gttttctgga 360
tattaactcc caggtttcca gaggaacatg gaggatttcc aacttaacct natggtttcc 420
gtagggtcag gaactgcagn tctattaggc ctttagngg cagtctggga tttgacntgg 480
ccagcattag ggggtgtattc acaantttgc tgnaggcctt ttgaccctat 530

```

<210> 79

<211> 495

<212> ADN

<213> Homo sapiens

<400> 79

```

ctgtatttgt atagagcaga aaatggatag tcagataaag aagcagactt tcacaatate 60
atatgttaat tttgtaatga gatagttgtt tagactagta aatagatcag gggttaaaggc 120
tttgattgaa tcagattatg aaagggtcca aacacaaaagg accgtccaag tctaactt 180
acagattcag ctggtaaaga aataatgaac taattttaga ttcagccaga ttattactta 240
tatacatagg tgaaggaaaa gagaagctaa aggtgtcagc cctatgagtc cttatccac 300
acacaaaaag ggtgggcact ggaaacaaan ggggtcagnt gggcttcagc nangngttgt 360
ggatacaccc tnttgctggc cngtcnattc cagactggca gctaantggc tatagagctg 420
cagttctgcc ctacggngnc ctagggggtta agttgggagt ctcntgttcc cctgggcnng 480
gggtttatat cngtt 495

```

<210> 80

<211> 594

<212> ADN

<213> Homo sapiens

<400> 80

```

catttaagggt tagaatctaa atcctttggc atagcactta agaccatttc caagggagtg 60
gctctaccat ggagccctag agactctaca tgtgtctaca tgggatgcac aggcaaagct 120
tggaactgcca tatgacctaa atctgggtgg aatgccctgt gatcctccta ggaacatggc 180
aagggaggaa atcttcagag aaggcttcgc gaggttggtt tctcatctgc ctctatcttg 240
tgggatgctg ccatgaggcg gttttctgat ataactcca gggtccagag gaacatgaga 300
cttccaaactt aaccctatg cttccgtagg gcagaactgc agctctatag ccatttagct 360
gcagtctgga attgactggc cagcaatagg gtgtatcaca actcttgctg aagccatctg 420
accccttttg tttcagtggc atccttttgg tgtgtgggat aaggactcat agggctgaca 480
tcttttagctt ctcttttctt cactatgtat ataagtaata atctggctga atctaaaatt 540
agttcattat ttctttacca gctgaatctg taagtgttag acttggaagg tcct 594

```

<210> 81
 <211> 582
 <212> ADN
 <213> Homo sapiens

<400> 81
 gttttctgga agtagggaga aaatcagagg ctaaagagtt tgaaaatgtg cacagaacag 60
 tttagagctca aaatgccact cctaccttat ttgcccaggt gattatcaat tcaccacaat 120
 tcccacagga aatgagaggc attatctttt gccaaataga acttgatagg ctattaactt 180
 gggtagcctg ggctggagg aggatggaaa taagatagaa gactaaaaac aggataacta 240
 gtggaagtct gaataaaagg ggatgaactc caactcctat cccaccatgt ggcgtgctc 300
 tgctgtctcg ccagtaatgg tagccaggct tatatcccc atgtgacta gaacatcctt 360
 ctctggggaa attaaagagc agacctctgg agaaattaca cttctcaaag aaaaattaac 420
 aggactgac atttagggtc ccaaaggcaa aaagtaggtc acactcaaag catctaactc 480
 aaagaccctg gacttcagtt gtagaccctg agcagattag actttggatt gtctcccttg 540
 ggtaaaatac agtggtctcc ccttattcaa agtttcactt tc 582

<210> 82
 <211> 321
 <212> ADN
 <213> Homo sapiens

<400> 82
 agtaaggctt ggcagttgac cttgtttgtt ggagagaagg gataagattt taaagctaca 60
 tgtctgaaag aatgatgctg ctgattgaaa taaaggaaga aaggatgcat ttcgggctcc 120
 aacctgtcct aggaaggcct agacctcaaa caccaacacc tccatgcatt tcctctttgg 180
 ctactatgtc ttttccctga cttctgcctc tccagctctc tgggctgctg cttccacctg 240
 ttcacttgac ttagaccctc cctgctgggt ccttggtcac ctactcattt ggtgcttcgt 300
 ctgccatcag tacctccatt g 321

<210> 83
 <211> 545
 <212> ADN
 <213> Homo sapiens

<400> 83
 gtcaattaca tcagcacttg gactggacca gggaaaggaa tgattctgct tcctgggaat 60
 gtcagaagga cctgatgatt atatttggca aagccaggag gagtggcttt gaatgtcatt 120
 gctaagaatt acactttgag tagcatttct ggatgtctga gcttttcaaa tgatacttct 180
 tttctgctgt ggctttcctt tctgttggac tgggtcccag agggctcctt tgtttgtcct 240
 tgccctcgct tttatatcag ttcattgttt ctcttctgtc atcttccctc ccagcgctgt 300
 ttctccaccc cctcctgctg cactcacaac agcttccctc ctctgttta gaggtggaag 360
 catgtaagaa tgcgtttgag ggggatgctt gccaaaggac agcatattca acatctggtg 420
 tcaacaagggt aatgtttaac cttagactag ccaaactagt gatgacctgc ttccatgctg 480

catctgctgc tttttgtgtt gatgggactc agaaatcatg agaaaggtct tcagtgatcc 540
atgac 545

<210> 84
<211> 183
<212> ADN
<213> Homo sapiens

<400> 84
caccaaagct gaagtctctc acttaatcct aagcaagggg ctaggagagaa ctaagatata 60
cttcccatat caagggatat aggttactag aagggtggag gtctcttaac aggctcactc 120
acctttcaga gtttaaatta tggtcaggct taagaagctc tctcttgtga ctgagtgtat 180
tgc 183

<210> 85
<211> 843
<212> ADN
<213> Homo sapiens

<400> 85
ggaatggaca tgaagaggat ttcacagtat gtagagcaga aggcacgtga atgtgtttgc 60
tttggcttgg agcttattaa gttttgacta cggttaaaat aagtcaaata gtaagtggta 120
aaacacattt ttgttagtat tggaaactttc tggagaacat aagggtctatg agaatgcata 180
tatatatattt ttaacatttc ctatatatct aaggtaacca agcactgagt ctaatttacc 240
tattaaggga gactctttaa aatcaacttt ataactaatt catactataa gacagataat 300
agctaaagtt ttggaataat ttatattaaa agccgcaagt cttaaaaatc cctggatatg 360
acataaaaag gattttggct tcttttttga agtattttaa attaatcacc ttagctctac 420
catatactag atctgtgacc gctacacaaa ttgtttatca tctttgggtc tctattgcct 480
tctttataaa ataagtgtaa gttgttcagc ctgcctcaca gggctgttgt gaggaaataa 540
aatgaaatgg agtatgtgaa attgcttcag caacagcaaa gtgctacgta aatgtaaagt 600
gttgttttta gctaataatg gatttaagtg tttggataat ttagatgca tttactttga 660
taaagcgtgt gcttaaagtg gtatcaccag tgatttctaa catgatttta aaaaaataaa 720
accccaatta aaatgttctt taatattcat ttaatttgtg catgacttgt ggcctttttg 780
tattttctca agcctattac tctagagctg taaaagctct tgcacagcat tgttgtgtca 840
gtg 843

<210> 86
<211> 613
<212> ADN
<213> Homo sapiens

<400> 86
aggttatcca cttttgctga agatattttt tattgaatca aagattgagt tacaattata 60
cttttcttac ctaagtggat aaaatgtact tttgatgaat cagggaattt ttttaaagtt 120
ggagtttagt tctaaattga ctttacgtat tactgcagtt aattcctttt ttggctaggg 180

```

atggtttgat aaaccacaat tggctgatat tgaaaatgaa agaaacttaa aagggtgggat 240
ggatcatgat tactgtcgat aactgcagat aaatttgatt agagtaataa ttttgtcatt 300
taaaaacaca gttgtttata ctgcccatcc taggatgctc accttccaag attcaacgtg 360
gctaaaacat cttctggtaa attgtgcgtc catattcatt ttgtcagtag ccaggagaaa 420
tggggatggg ggaaatacga cttagtggag catagacatc cctgggccat cttttctgtc 480
tccagctgtt tcttggaacc tgctctctct cttgtctgtc cctgacgcag agaccgttgc 540
ctccccaca gccgtttgac tgaaggctgc tctggagacc tagagtaaaa cggctgatgg 600
aagttgtggg acc 613

```

<210> 87

<211> 705

<212> ADN

<213> Homo sapiens

<400> 87

```

gctaattgtt cttgctgggt aattattagc ataaatatca ggttcttaga ggtctttaag 60
agaatacttt agagctacca aaattagcag cagtgatgta atatcaccct gaaataccag 120
gagcagaggt gccccagagt gcttcacagc tatgatcagg aaacttgacc tgccccattg 180
ctacagcaac ggcataagga ctaggaaggt ccactagctg gggaggctt ggaaccctcc 240
gagccagaaa gaagcattcc agcatgattc ctcagatgca atgaggaagg tttacagaat 300
tttcacaggg caacagagat caagtggat gaagggtag gatttttaag ctcaagcttc 360
aaagataacc tacctttaca ttctcagtca gaatttgacc cagaaagagg aagtttcttc 420
aggaacagac cgtacgttga tgaagtagaa gtcatttcag taaaaacgga gtacagaact 480
tggtcataat atcttgcatt ttatagattt attaaagatt agtttcaagt tcacattcgc 540
tattcagttg taaaccgaat ggatgggagg ggagaaaata caagctctcc acacaggtat 600
gtcctctctt tttctgagag agaaggcatg ggattttcag cataaattcc atgttatgtg 660
agtgtgtgtt gatttctgaa gttcctatca atatctgttc ctgca 705

```

<210> 88

<211> 459

<212> ADN

<213> Homo sapiens

<400> 88

```

ctcgttgctc agtttagggc actactctta aaaaaggaaa gttaacaaac tggaatagag 60
tcagagataa ctttgagaaa accgatgtca ttaaactggg gtctctggac ctgaggtttg 120
cactcacatt tccatctggc ggccccataa gcaatctgtc ctacagataa ctctctctac 180
acaaaactta gtctcttttc agctcagctc tctcactctc aattatatct ccttacttcc 240
atatggcact gttgtacact catttactca gagccagaaa cgtcagcgtc atcttggatt 300
tttcttatgc tctttctctc tctagtcata tgccagactt taaactctgc ttgaaagctt 360
tctcataagc tctttccttt tccctttcta ctgctttgca tttgctactt aaccttttc 420
ttcaggctgt ttgctttcca gtccatcggt cgctctgct 459

```

<210> 89

<211> 775

<212> ADN

<213> Homo sapiens

<400> 89

```
tctatccgctc cgcgtcagcg ccttgccacc ctcatctcca atatgcctgg tccgaccccc 60
agtggcacta acgtgggatc ctcagggcgc tctcccagca aagcagtggc cgccccggcg 120
gcgggatcca ctgtccggca gaggttaagga accctgcagt tcgttcgctt ccagactcgg 180
agataggacc cagaacctcg ctgattcttg ggtggagacc ctagcatgtg aagattgaca 240
aaggcaaaat gagcttctag tgacgtggcc gtgggagtag ttaaaggcct tttgggagga 300
aggcgacatt ttttttctcg ttgctcagtt tagggcacta ctcttaaaaa aggaaagtta 360
acaaactgga atagagtcag agataacttt gagaaaaccg atgtcattaa actggtgtct 420
ctggacctga ggtttgact cacatttcca tctggcggcc ccataagcaa tctgtcctac 480
agataactcg tcctacacaa aacttagtct cttttcagct cagctctctc actctcaatt 540
atatctcctt acttccatat ggcactgttg tacactcatt tactcagagc cagaaacgtc 600
agcgtcatct tggatttttc ttatgctctt tctctctcta gtcatatgcc agactttaaa 660
ctctgcttga aagctttctc ataagctctt tccttttccc tttctactgc tttgcatttg 720
ctacttaacc cttttcttca ggctgtttgc tttccagtc atcgttcgct ctgct 775
```

<210> 90

<211> 727

<212> ADN

<213> Homo sapiens

<400> 90

```
gaatgctgtt acctgaaccc cttacttcga aggatcataa gattcacagg ggtgtttgca 60
tttggacttt ttgctactga cttttttgta aacgccggac aagtggtcac tgggcactta 120
acgccatact tcctgactgt gtgcaagcca aactacacca gtgcagactg ccaagcgcac 180
caccagttta taaacaatgg gaacatttgt actggggacc tgggaagtga agaaaaggct 240
cggagatcct ttccctccaa acacgctgct ctgagcattt actccgcctt atatgccacg 300
atgtatatta caagcacaat caagacgaag agcagtcgac tggccaagcc ggtgctgtgc 360
ctcggaactc tctgcacagc cttcctgaca ggcctcaacc gggctctctga gtatcggaac 420
cactgctcgg acgtgattgc tggtttcac ctagggcactg cagtggccct gtttctggga 480
atgtgtgtgg ttcataactt taaaggaacg caaggatctc cttccaaacc caagcctgag 540
gatccccgtg gagtaccctt aatggccttc ccaaggatag aaagccctct ggaaacctta 600
agtgcacaga atcactctgc gtccatgacc gaagttacct gagacgactg atgtgtcaca 660
agctgttttt taaaatcatc ttccaattct atacttcaaa acacacagtt gctcaatgtc 720
aaactgt 727
```

<210> 91

<211> 478

<212> ADN

<213> Homo sapiens

<400> 91

```
cagtttttat cacagggtgca agccgtggca ttggcaaagc tattgcattg aaagcagcaa 60
aggatggagc aaatattggt attgctgcaa agaccgccca gccacatcca aaacttctag 120
```

```

gcacaatcta tactgctgct gaagaaattg aagcagttgg aggaaaggcc ttgccatgta 180
ttgttgatgt gagagatgaa cagcagatca gtgctgcagt ggagaaaggcc atcaagaaat 240
ttggaggaat tgatattctg gtaaataatg ccagtgccat tagtttgacc aatacattgg 300
acacacctac caagagattg gatctgatga tgaacgtgaa caccagaggc acctaccttg 360
catctaaagc atgtattcct tatttgaaaa agagcaaagt tgctcatatc ctcaatatca 420
gtccaccact gaacctaaat ccagtttggt tcaaacagca ctgtgcttat accattgc 478

```

<210> 92

<211> 679

<212> ADN

<213> Homo sapiens

<400> 92

```

gccaggaaac agcaaatcca atagatccga ctttatcagg ctgggctctc gcttcagatt 60
cagtgggagg acagaatata aagctacaca tggctccagg ttacgaagaa ccagcacctt 120
tgagaggaag cctagtaaac gttatccatc ccggagacat tcaacgttca aagcaagcaa 180
cccagtgata gcagcccagc tctgctctaa acaaatcca gaagtccata attaccaggc 240
tcaatatcat cctaataatcc atcccagcca gcccgggtgg catcctcact ctccaaatgt 300
caggccatcc ttccaggatg acaggctcga ttggaaagca tcggccagtg gagatgacag 360
ccatthttgat tatgtccacg accagaacca gaagaactta ggagggatgc aaagtatgat 420
gtatcgagat aaactcatga ctgcactttg agagactgaa gcatctctct tccattcacc 480
ttcatagttt cattgcattc catgaaaagt gtcttggcct cagatggatg gatgtgtttg 540
gacgagtgtc tttaaggagt agtccctgaaa ggtgtttttg gtgtccatgt aaatatttga 600
agataaaacc actatagctt gtcataatth actgttgact gcattctcat taaaatgaag 660
gtaaaggctc aggaatcat

```

679

<210> 93

<211> 691

<212> ADN

<213> Homo sapiens

<400> 93

```

ctgaggatga ttatggtgat gacccttcga ccaactcggt tgagaaagag aagcaggact 60
atgtctattg cctggagtca tccctgcaga cctacaaccc agactacgtc ctgatggtag 120
aagacgatgc tgtaccagaa gagecagatc tcccagtcct ggagcacctt ctgagggtc 180
gcttctctga gccacatctc agagatgccc tttatctcaa gctgtatcac cccgagaggc 240
tocagacta catcaatcca gagcccatgc ggatcctgga atgggttggg gtaggcatgt 300
tgctggggcc cttactaacc tggatataca tgaggtttgc cagccgccc gggtttagct 360
ggcctgtaat gctctctctc tccctgtata gcatgggtct ggtggagctg gtgggtcggc 420
actatttctt ggaactgcgg cggctgagtc cttccctgta cagtgtgggt cctgcctctc 480
agtgttgca cccagccatg ctcttcccgg cactgcgggc ccgcccggac ctcacctacc 540
tgtcccaagt gtactgccac aagggttttg gcaaggacat ggcaactgtac tcgctgttga 600
gggccaaggg agagagggcc tatgtagtgg agccgaacct cgtgaaacac atcgggtctc 660
tctccagttc ccgttacaac tttcatccca g

```

691

<210> 94
<211> 451
<212> ADN
<213> Homo sapiens

<400> 94
caactttctc cattcccagc tgggccccat tectggattt aagatgggtg ctatccctga 60
ggagtcacca taaggagaaa actcaggaat tctgagtctt ccctgctaca ggaccagttc 120
tgtgcaatga acttgagact cctgatgtac actgtgatat tgaccgaagg ctacatacag 180
atctgtgaat cttggctggg acttcctctg agtgatgcct gagggtcagc tcctctagac 240
attgactgca agagaatctc tgcaacctcc tatataaaag catttctgtt aattcattca 300
gaatccattc tttaacaatat gcagtgagat gggcttaagt ttgggctaga gtttgacttt 360
atgaaggagg tcattgaaaa agagaacagt gacgtaggca aatgtttcaa gcactttaga 420
aacagtactt ttcctataat tagttgatat a 451

<210> 95
<211> 662
<212> ADN
<213> Homo sapiens

<400> 95
ggacctttgc tgctgttggc cactccgtca ccttttggaa cacaagtata tcaagatagt 60
ggctactgat gttcaagtgg gattgaagtg tgataaatgg atatattttg ttgtttgctg 120
gggtgttcat ggagatgtta agagattgag gccctgggct gagggtatat aatgtatgtc 180
aggtaaagt tgaagactgc caaggagcag attttctccc tggaaatgtg aaaactgaac 240
ctataactct gataaggact tgagatgtgt agaaacgttg ggttatggaa gactagtctc 300
ttccataacc ctgaatttga gaccttaatg ctaagtgtag attattgagg tttgttagtg 360
aggaaaagaa taagagttca gaagcctttg ttatcagata gcgaaatcag ggcctagtga 420
ggagcacagg tcgactacat aatggagtcc attggcgaac cctattgcaa tttggtccaa 480
ctatatcttc tgggtgaagga aattaatgat gtaagaaaat gcaagaggct caacttctct 540
tccaaaaatc ttctggcttc tgaactcttc ctctgcctct cttaaataa ataacacaga 600
atttcaagtg gtaggagact tattaagcca gtcaccaagc ttggtctgtc agcctgtctt 660
ct 662

<210> 96
<211> 766
<212> ADN
<213> Homo sapiens

<400> 96
ctatcctttc agcccaccaa cttagcggca gcactaggga ttcattataa ggtaaactctg 60
gtttacataa agacctgaag gaggcctgta tttgaagctc acacttggtg ttggtatctc 120
tcatttttac tgagccagtg tggaaatacca ctgtatgtac tcatataagc ccttgacttt 180
tactgtcat caggattgga atattactct agcagtcttc acacataggc aagttacagt 240
ccttttaaaa agnatctcat ttcctataa tggaaacctaa tagccaactt tttcatagaa 300
attgctagaa gagtttgatc aactataaat gataaagtgt ttataagcat agtcagtgtg 360

```

acacagaaac caatcttaaa attgaattta atgttttatc atatcagatt aaatattttc 420
tccatgtctt atttttactg caacaagtta gaaagtggga acactttgat taatgtctta 480
aaatttggtg gccctcattt ggataaaggg agcaatccta aggacttttt ttttttttaa 540
cataatctga gaatttctct gtagagcaga gactttcaaa ctttttggct gtaaccacaca 600
gtaaaaaacg cttttatata aaaccttaga atatgtttta tgaacaatac ttaccattct 660
gatgcttttt attgtttcag tttttaaaat atgccagtgt caaccacta aattgatata 720
taccaatggg ttgcaaccct tagcttgaaa aaaacaccct cacaga 766

```

<210> 97

<211> 584

<212> ADN

<213> Homo sapiens

<400> 97

```

agttgatcga ttcattgtct attgatagaa tcttgaccag aagaaatttt gctcttttta 60
tatagtttca agaaatgtgt ttttaaattt ttattaatgc acttgaacaa ctttgcagga 120
ataaagcaac cccctaacca caaaatatcc ctctaaatta gttccctagc tttctcaatg 180
aatacacaca tatttttaca tagctatgat cgttggtgtac attctccttt gttttacttc 240
tcggcctaac acttgtctcc tcttgtcaac acagattcta ctctcaccaa tttaaatgtc 300
tttatatcca tgtaacatgg gtaacctcac ttcaccccat tattagatat ttgagttata 360
tctaattttt cactcttata aatagtgtct ctatgaatgt ctgtaaaaaa aaaaaaactg 420
ctccttcttt tggattatcc ccttaggaat atctccaaag agggattaca aggtcaaaga 480
gcatgaagta ttttatagct cttgttttat attgccagat tgctttctag aaagatccaa 540
tctttggggt ggaaggacct taaagggtcat ctagtttagc ctcc 584

```

<210> 98

<211> 1251

<212> ADN

<213> Homo sapiens

<400> 98

```

gtgtgtagca gcggtgtccaa ggcaagaact gaaaactctg agtgatgttc agaagtttgc 60
agagataaat gggtcagccc tatgtagcta cancctaaag cttctgaat acactacatc 120
tccaaaatct tctgttctct gccccaaact accagttcca gcgagtgcac ctattccatt 180
cttccatcgc tgtgtcctct tgaacatttc ctgctatgcc aagtttgagc aggccctgat 240
cacctttgtc agtgacaata gtgtcttaca caggctgatt agtggagtaa tgaccagcaa 300
agaaattata ttgggacttt gcttgttata actagttcta tccatgattt tgatgggtgat 360
aatcaggtat atatcaagag tacttgtgtg gatcttaacg attctggtea tactcggttc 420
acttggaggc acaggtgtac tatgggtggc gtatgcaaag caaagaaggt ctcccaaaga 480
aactgttact cctgagcagc ttcagatagc tgaagacaat cttcgggccc tcttcattta 540
tgccatttca gctacagtgt tcacagtgat ctattcctg ataagtgttg ttatgcgcaa 600
acgtgttgct cttaccatcg ccttgttcca cgtagctggc aaggtcttca ttcacttgcc 660
actgctagtc ttccaaccct tctggacttt ctttgcctct gtcttgtttt ggggtgactg 720
gatcatgaca cttctttttc ttggcactac cggcagtcct gttcagaatg agcaaggctt 780
tgtggagttc aaaatttctg ggcctctgca gtacatgttg tggtagcatg tgggtgggct 840
gatttggatc agtgaattta ttctagcatg tcagcagatg acagtggcag gagctgtggt 900

```



```

aacatactat tttactaggg ataaaaggaa tttgccattt acacctattt tggcatcagt 960
aaatcgcctt attcgttacc acctaggtag ggtggcaaaa ggatctttca ttatcacatt 1020
agtcaaaaatt ccgcgaatga tccttatgta tattcacagt cagctcaaag gaaaggaaaa 1080
tgcttggtgca cgatgtgtgc tgaaatcttg catttggtgc ctttggtgct ttgaaaagtg 1140
cctaaattat ttaaatacaga atgcatacac agccacagct atcaacagca ccaacttctg 1200
cacctcagca aaggatgcct ttgtcattct ggtggagaat gctttgcgag t 1251

```

<210> 99

<211> 710

<212> ADN

<213> Homo sapiens

<400> 99

```

tttgatcctc cagtgtgact gttgtttttg tttgggggtg ggttgggggt ttttgctttt 60
tattcctgaa gcttaccaga tatgaatggc taatactcca ttgttctgct tgttgtaatg 120
gtgaatgctt taagaaaaaa aagtgtaat tgctaagaat aattcatgat ctgtttatgc 180
gataactcct ttttgttaca atttttttta aaaaagctat ttttgtaat gtaaagtaaa 240
tatttcagag caaatTTTT aaacttattg cactaaatac aggctctgta caaaaaaaaa 300
aaaaaaaaaa aaaagcctca gcattttatc attccatgga aggagaatct tttgaaagaa 360
agcattgcct cctaccagaa ctgacagtg aattagatcg gtattatgga aatgcataca 420
agtaatgtca ctagggtta ataagcagcc gtttgctaatt gtgcttctct tcaaagggtt 480
ggacctttaa attgctgcaa aaggtaaatt gtattttttt ttaagtattg gtgttcttta 540
ctctagctag gctaaaaatt gctaaatgcc ttgggttctt ttaaaagttc atgtaatat 600
tctgattttt cagaatattt gcaataagag tctggatttt aaaaaacaca tgcatacaca 660
caattaagag ctcatgtctt agcaagatct gggaaaccaa cattgcgaga 710

```

<210> 100

<211> 580

<212> ADN

<213> Homo sapiens

<400> 100

```

aggttcctgt agtctgtgct cagaacttgg tttttggccc ctattgtttt tgcctatttt 60
gattttcaga gatgatcaca tggggacagt taacttttct tctgctgtgt tgccttaatg 120
ctactagatt gtgttggtgt gttggagttt tctgacttct tccctataaa aagatactga 180
gagctccata atgaaagaag ttgttatact ttctcagaat attctggacc actgaatgca 240
cttctaatag agctttaatc taaagaagtt agttcagtg ttattaactg attttattac 300
aggagaaaaa aactttaaca aaaaggcagg gagaaaagtg tgaagggtcat caagcaaaat 360
gacaggggct tcaaaaaaca accaaagaca aaaccctatc ttctgaagac caaagggtcca 420
actttactta ctggctggca cagcctttct gaactccttg agtttagaat agagctccta 480
gaataataag gcggccaaat tttaaagatc agtcaatata gtagggacct gctattgatc 540
tctcaggcac tgagttcttc acatccagt tcaagcccag 580

```

<210> 101

<211> 223

<212> ADN

<213> Homo sapiens

<400> 101

atgacagaga ttgacctcc aggggggtcta atttgaagag gaaagtaatt taaatgtata 60
ggtttatata atagcattgt atttttctct attttattca ttctttcatt ctcttcgatg 120
actcccagca cctgcgcgtc catcttcatt cttageccaat gacctttcct tttacttaac 180
agagaaaata acaatcagaa gagtcttcat aagtccccac ccc 223

<210> 102

<211> 19

<212> ADN

<213> Homo sapiens

<400> 102

ctcctttccc accagacac 19

<210> 103

<211> 20

<212> ADN

<213> Homo sapiens

<400> 103

cattccaagg ttccctttttt 20

<210> 104

<211> 22

<212> ADN

<213> Homo sapiens

<400> 104

agctcctctt gcctaagtgg tc 22

<210> 105

<211> 22

<212> ADN

<213> Homo sapiens

<400> 105

gtttgtgtgg ttctgttttc ca 22

<210> 106

<211> 19

<212> ADN
<213> Homo sapiens

<400> 106
aaccctgtgc tgatctccc

19

<210> 107
<211> 18
<212> ADN
<213> Homo sapiens

<400> 107
gtgcttcaaa ggcttcaa

18

<210> 108
<211> 17
<212> ADN
<213> Homo sapiens

<400> 108
tcagtggaaa gaaatgt

17

<210> 109
<211> 18
<212> ADN
<213> Homo sapiens

<400> 109
atggggacga tttcagtg

18

<210> 110
<211> 21
<212> ADN
<213> Homo sapiens

<400> 110
ctgctgtcta ttgggaaagt t

21

<210> 111
<211> 20
<212> ADN
<213> Homo sapiens

<400> 111
agaagcagtg gacaggggag

20

<210> 112
<211> 22
<212> ADN
<213> Homo sapiens

<400> 112
ctggttgaac ctgtggattg tg

22

<210> 113
<211> 22
<212> ADN
<213> Homo sapiens

<400> 113
tggagagcag cttccaaaaa tc

22

<210> 114
<211> 18
<212> ADN
<213> Homo sapiens

<400> 114
caatgcacct gtcagggg

18

<210> 115
<211> 22
<212> ADN
<213> Homo sapiens

<400> 115
tgttggacct ggtactctcc ac

22

<210> 116
<211> 18
<212> ADN
<213> Homo sapiens

<400> 116
gggcaaattg ggaaacat

18

<210> 117
<211> 19
<212> ADN
<213> Homo sapiens

<400> 117
gggaagacag gggaaagag

19

<210> 118
<211> 22
<212> ADN
<213> Homo sapiens

<400> 118
ccattgactc ctccttcttt ct

22

<210> 119
<211> 19
<212> ADN
<213> Homo sapiens

<400> 119
agtaacaaca taccctgcc

19

<210> 120
<211> 21
<212> ADN
<213> Homo sapiens

<400> 120
cagacatgac tgaattggtg g

21

<210> 121
<211> 20
<212> ADN
<213> Homo sapiens

<400> 121
aaacaccccc atcccctcac

20

<210> 122
<211> 19

<212> ADN
<213> Homo sapiens

<400> 122
ggggctgaca gcagggttt

19

<210> 123
<211> 20
<212> ADN
<213> Homo sapiens

<400> 123
aaacccacc ccagttcctt

20

<210> 124
<211> 21
<212> ADN
<213> Homo sapiens

<400> 124
gaagtgaacg ggctgttttg c

21

<210> 125
<211> 19
<212> ADN
<213> Homo sapiens

<400> 125
ggagctgaaa ggtggaatg

19

<210> 126
<211> 22
<212> ADN
<213> Homo sapiens

<400> 126
agtgaacaag gaagttaagg gc

22

<210> 127
<211> 21
<212> ADN
<213> Homo sapiens

<400> 127

cccggattca tttaagggtta g

21

<210> 128

<211> 18

<212> ADN

<213> Homo sapiens

<400> 128

cacaaaggac cgtccaag

18

<210> 129

<211> 461

<212> PRT

<213> Homo sapiens

<400> 129

Met	Asp	Ser	Gln	Lys	Tyr	Cys	Phe	Lys	Glu	Asn	Glu	Asn	Val	Thr	Val
1				5					10					15	

Asp	Lys	Ala	Cys	Phe	Leu	Ile	Ser	Asn	Ile	Thr	Ile	Gly	Pro	Glu	Ser
		20						25					30		

Ile	Asn	Leu	Gln	Gln	Glu	Ala	Leu	Gln	Arg	Ile	Ile	Ser	Thr	Leu	Ala
	35						40					45			

Asn	Lys	Asn	Asp	Glu	Ile	Gln	Asn	Phe	Ile	Asp	Thr	Leu	His	His	Thr
	50					55					60				

Leu	Lys	Gly	Val	Gln	Glu	Asn	Ser	Ser	Asn	Ile	Leu	Ser	Glu	Leu	Asp
65				70						75					80

Glu	Glu	Phe	Asp	Ser	Leu	Tyr	Ser	Ile	Leu	Asp	Glu	Val	Lys	Glu	Ser
				85					90					95	

Met	Ile	Asn	Cys	Ile	Lys	Gln	Glu	Gln	Ala	Arg	Lys	Ser	Gln	Glu	Leu
			100					105					110		

Gln	Ser	Gln	Ile	Ser	Gln	Cys	Asn	Asn	Ala	Leu	Glu	Asn	Ser	Glu	Glu
		115					120					125			

Leu	Leu	Glu	Phe	Ala	Thr	Arg	Ser	Leu	Asp	Ile	Lys	Glu	Pro	Glu	Glu
	130					135					140				

Phe	Ser	Lys	Ala	Ala	Arg	Gln	Ile	Lys	Asp	Arg	Val	Thr	Met	Ala	Ser
145					150					155					160

Ala Phe Arg Leu Ser Leu Lys Pro Lys Val Ser Asp Asn Met Thr His
 165 170 175
 Leu Met Val Asp Phe Ser Gln Glu Arg Gln Met Leu Gln Thr Leu Lys
 180 185 190
 Phe Leu Pro Val Pro Lys Ala Pro Glu Ile Asp Pro Val Glu Cys Leu
 195 200 205
 Val Ala Asp Asn Ser Val Thr Val Ala Trp Arg Met Pro Glu Glu Asp
 210 215 220
 Asn Lys Ile Asp His Phe Ile Leu Glu His Arg Lys Thr Asn Phe Asp
 225 230 235 240
 Gly Leu Pro Arg Val Lys Asp Glu Arg Cys Trp Glu Ile Ile Asp Asn
 245 250 255
 Ile Lys Gly Thr Glu Tyr Thr Leu Ser Gly Leu Lys Phe Asp Ser Lys
 260 265 270
 Tyr Met Asn Phe Arg Val Arg Ala Cys Asn Lys Ala Val Ala Gly Glu
 275 280 285
 Tyr Ser Asp Pro Val Thr Leu Glu Thr Lys Ala Leu Asn Phe Asn Leu
 290 295 300
 Asp Asn Ser Ser Ser His Leu Asn Leu Lys Val Glu Asp Thr Cys Val
 305 310 315 320
 Glu Trp Asp Pro Thr Gly Gly Lys Gly Gln Glu Ser Lys Ile Lys Gly
 325 330 335
 Lys Glu Asn Lys Gly Ser Val His Val Thr Ser Leu Lys Lys His Thr
 340 345 350
 Arg Ser Gly Thr Pro Ser Pro Lys Arg Thr Ser Val Gly Ser Arg Pro
 355 360 365
 Pro Ala Val Arg Gly Ser Arg Asp Arg Phe Thr Gly Glu Ser Tyr Thr
 370 375 380
 Val Leu Gly Asp Thr Ala Ile Glu Ser Gly Gln His Tyr Trp Glu Val
 385 390 395 400
 Lys Ala Gln Lys Asp Cys Lys Ser Tyr Ser Val Gly Val Ala Tyr Lys
 405 410 415

Thr Leu Gly Lys Phe Asp Gln Leu Gly Lys Thr Asn Thr Ser Trp Cys
 420 425 430

Ile His Val Asn Asn Trp Leu Gln Asn Thr Phe Ala Ala Lys His Asn
 435 440 445

Asn Lys Val Lys Ala Leu Asp Val Thr Val Leu Lys Lys
 450 455 460

<210> 130

<211> 32

<212> PRT

<213> Homo sapiens

<400> 130

Ile Lys Asn Ile Ile Leu Ser Val Ser Leu Trp Ser Lys Met Phe Met
 1 5 10 15

Val Glu Thr Gly Asp Leu Ala Ala Ser Leu Thr Thr Phe Ser Lys Cys
 20 25 30

<210> 131

<211> 183

<212> PRT

<213> Homo sapiens

<400> 131

Asn Ile Ile Gly Gln Ser Val Glu Glu Ala Ile Arg Gly Val Phe Asp
 1 5 10 15

Ala Ser Leu Lys Met Ala Gly Phe Tyr Gly Leu Tyr Thr Trp Leu Thr
 20 25 30

His Thr Met Phe Gly Ile Asn Ile Val Phe Ile Pro Ser Ala Leu Ala
 35 40 45

Ala Ile Leu Gly Ala Val Pro Phe Leu Gly Thr Tyr Trp Ala Ala Val
 50 55 60

Pro Ala Val Leu Asp Leu Trp Leu Thr Gln Gly Leu Gly Cys Lys Ala

[illegible]

Pro His Asp Leu Val Ser Gln Gly Tyr Ser Tyr Leu Pro Val Gln Leu
 100 105 110
 Leu Ala Ser Gly Met Lys Glu Val Thr Arg Thr Trp Lys Ile Val Gly
 115 120 125
 Gly Val Thr His Ala Asn Ser Tyr Tyr Lys Asn Gly Trp Ile Val Met
 130 135 140
 Ile Ala Ile Gly Trp Ala Arg Gly Ala Gly Gly Thr Ile Ile Thr Asn
 145 150 155 160
 Phe Glu Arg Leu Val Lys Gly Asp Trp Lys Pro Glu Gly Asp Glu Trp
 165 170 175
 Leu Lys Met Ser Tyr Pro Ala Lys Val Thr Leu Leu Gly Ser Val Ile
 180 185 190
 Phe Thr Phe Gln His Thr Gln His Leu Ala Ile Ser Lys His Asn Leu
 195 200 205
 Met Phe Leu Tyr Thr Ile Phe Ile Val Ala Thr Lys Ile Thr Met Met
 210 215 220
 Thr Thr Gln Thr Ser Thr Met Thr Phe Ala Pro Phe Glu Asp Thr Leu
 225 230 235 240
 Ser Trp Met Leu Phe Gly Trp Gln Gln Pro Phe Ser Ser Cys Glu Lys
 245 250 255
 Lys Ser Glu Ala Lys Ser Pro Ser Asn Gly Val Gly Ser Leu Ala Ser
 260 265 270
 Lys Pro Val Asp Val Ala Ser Asp Asn Val Lys Lys Lys His Thr Lys
 275 280 285
 Lys Asn Glu
 290

<210> 133

<211> 48

<212> PRT

<213> Homo sapiens

<400> 133

Ser Ile Arg Pro Arg Gln Arg Leu Ala Thr Leu Ile Ser Asn Met Pro
 1 5 10 15
 Gly Pro Thr Pro Ser Gly Thr Asn Val Gly Ser Ser Gly Arg Ser Pro
 20 25 30
 Ser Lys Ala Val Ala Ala Arg Ala Ala Gly Ser Thr Val Arg Gln Arg
 35 40 45

<210> 134
 <211> 22
 <212> PRT
 <213> Homo sapiens

<400> 134
 Asn Lys Thr Gln Arg Pro Cys Ala Gln Gln His Ser Arg Met Pro Pro
 1 5 10 15

Ile Pro Arg Phe Pro Phe
 20

<210> 135
 <211> 95
 <212> PRT
 <213> Homo sapiens

<400> 135
 Met Trp Asp Phe Arg Ala Arg Ile Phe Asn Ser Gln Gln Tyr Trp Lys
 1 5 10 15

Leu Arg Leu Gln Asn Pro Gly Met Glu Gly Arg Glu Phe Glu Leu Cys
 20 25 30

Ser Leu His Arg Met Gly Ser Gln Glu Asn Ala Asn Val Lys Leu Ser
 35 40 45

Tyr Arg Glu Val Ser Ile Asn Leu His Arg Thr Pro Thr Lys Asp Thr
 50 55 60

Leu Arg Glu Lys Val Cys Ser Pro Lys Tyr Thr Ser Leu Phe Asn Leu
 65 70 75 80

Leu Pro Lys Ile Arg Glu Asn Ala Asp Thr Ala Gly Lys Arg Ser
 85 90 95

<210> 136
 <211> 108
 <212> PRT
 <213> Homo sapiens

<400> 136
 Leu Lys Thr His Phe Pro Pro Cys Pro Arg His Lys Phe Ile Leu Met
 1 5 10 15

Leu Ile Leu Gln Ser Leu Arg Tyr His Ala Ser Tyr Arg Asp Ile Leu
 20 25 30

Asp Asp Leu Leu His Gln Asp Arg Ile Gly Gln His Met Ser Pro Glu
 35 40 45

Gly Cys Met Leu Lys Gln Gln Ser Lys Gln Leu Ala His Asn Arg Gln
 50 55 60

Leu Leu Arg Pro Asn Ser Val His Pro Gln Ala Leu Gln His Gln Pro
 65 70 75 80

Glu Asp Ser Arg Gln Glu Ser Gln Asp Gly Asp Thr Asp Pro Pro Gln
 85 90 95

Leu Val Pro Thr Ile Thr Glu Thr Glu Met Ser Gln
 100 105

<210> 137
 <211> 213
 <212> PRT
 <213> Homo sapiens

<400> 137
 Glu Cys Cys Tyr Leu Asn Pro Leu Leu Arg Arg Ile Ile Arg Phe Thr
 1 5 10 15

Gly Val Phe Ala Phe Gly Leu Phe Ala Thr Asp Ile Phe Val Asn Ala
 20 25 30

Gly Gln Val Val Thr Gly His Leu Thr Pro Tyr Phe Leu Thr Val Cys

35 40 45
 Lys Pro Asn Tyr Thr Ser Ala Asp Cys Gln Ala His His Gln Phe Ile
 50 55 60
 Asn Asn Gly Asn Ile Cys Thr Gly Asp Leu Glu Val Ile Glu Lys Ala
 65 70 75 80
 Arg Arg Ser Phe Pro Ser Lys His Ala Ala Leu Ser Ile Tyr Ser Ala
 85 90 95
 Leu Tyr Ala Thr Met Tyr Ile Thr Ser Thr Ile Lys Thr Lys Ser Ser
 100 105 110
 Arg Leu Ala Lys Pro Val Leu Cys Leu Gly Thr Leu Cys Thr Ala Phe
 115 120 125
 Leu Thr Gly Leu Asn Arg Val Ser Glu Tyr Arg Asn His Cys Ser Asp
 130 135 140
 Val Ile Ala Gly Phe Ile Leu Gly Thr Ala Val Ala Leu Phe Leu Gly
 145 150 155 160
 Met Cys Val Val His Asn Phe Lys Gly Thr Gln Gly Ser Pro Ser Lys
 165 170 175
 Pro Lys Pro Glu Asp Pro Arg Gly Val Pro Leu Met Ala Phe Pro Arg
 180 185 190
 Ile Glu Ser Pro Leu Glu Thr Leu Ser Ala Gln Asn His Ser Ala Ser
 195 200 205
 Met Thr Glu Val Thr
 210

<210> 138
 <211> 325
 <212> PRT
 <213> Homo sapiens

<400> 138
 Met Ala Val Gly Asn Asn Thr Gln Arg Ser Tyr Ser Ile Ile Pro Cys
 1 5 10 15
 Phe Ile Phe Val Glu Leu Val Ile Met Ala Gly Thr Val Leu Leu Ala
 20 25 30

Tyr Tyr Phe Glu Cys Thr Asp Thr Phe Gln Val His Ile Gln Gly Phe
 35 40 45
 Phe Cys Gln Asp Gly Asp Leu Met Lys Pro Tyr Pro Gly Thr Glu Glu
 50 55 60
 Glu Ser Phe Ile Thr Pro Leu Val Leu Tyr Cys Val Leu Ala Ala Thr
 65 70 75 80
 Pro Thr Ala Ile Ile Phe Ile Gly Glu Ile Ser Met Tyr Phe Ile Lys
 85 90 95
 Ser Thr Arg Glu Ser Leu Ile Ala Gln Glu Lys Thr Ile Leu Thr Gly
 100 105 110
 Glu Cys Cys Tyr Leu Asn Pro Leu Leu Arg Arg Ile Ile Arg Phe Thr
 115 120 125
 Gly Val Phe Ala Phe Gly Leu Phe Ala Thr Asp Ile Phe Val Asn Ala
 130 135 140
 Gly Gln Val Val Thr Gly His Leu Thr Pro Tyr Phe Leu Thr Val Cys
 145 150 155 160
 Lys Pro Asn Tyr Thr Ser Ala Asp Cys Gln Ala His His Gln Phe Ile
 165 170 175
 Asn Asn Gly Asn Ile Cys Thr Gly Asp Leu Glu Val Ile Glu Lys Ala
 180 185 190
 Arg Arg Ser Phe Pro Ser Lys His Ala Ala Leu Ser Ile Tyr Ser Ala
 195 200 205
 Leu Tyr Ala Thr Met Tyr Ile Thr Ser Thr Ile Lys Thr Lys Ser Ser
 210 215 220
 Arg Leu Ala Lys Pro Val Leu Cys Leu Gly Thr Leu Cys Thr Ala Phe
 225 230 235 240
 Leu Thr Gly Leu Asn Arg Val Ser Glu Tyr Arg Asn His Cys Ser Asp
 245 250 255
 Val Ile Ala Gly Phe Ile Leu Gly Thr Ala Val Ala Leu Phe Leu Gly
 260 265 270
 Met Cys Val Val His Asn Phe Lys Gly Thr Gln Gly Ser Pro Ser Lys
 275 280 285

Pro Lys Pro Glu Asp Pro Arg Gly Val Pro Leu Met Ala Phe Pro Arg
 290 295 300

Ile Glu Ser Pro Leu Glu Thr Leu Ser Ala Gln Asn His Ser Ala Ser
 305 310 315 320

Met Thr Glu Val Thr
 325

<210> 139

<211> 406

<212> PRT

<213> Homo sapiens

<400> 139

Val Phe Ile Thr Gly Ala Ser Arg Gly Ile Gly Lys Ala Ile Ala Leu
 1 5 10 15

Lys Ala Ala Lys Asp Gly Ala Asn Ile Val Ile Ala Ala Lys Thr Ala
 20 25 30

Gln Pro His Pro Lys Leu Leu Gly Thr Ile Tyr Thr Ala Ala Glu Glu
 35 40 45

Ile Glu Ala Val Gly Gly Lys Ala Leu Pro Cys Ile Val Asp Val Arg
 50 55 60

Asp Glu Gln Gln Ile Ser Ala Ala Val Glu Lys Ala Ile Lys Lys Phe
 65 70 75 80

Gly Gly Ile Asp Ile Leu Val Asn Asn Ala Ser Ala Ile Ser Leu Thr
 85 90 95

Asn Thr Leu Asp Thr Pro Thr Lys Arg Leu Asp Leu Met Met Asn Val
 100 105 110

Asn Thr Arg Gly Thr Tyr Leu Ala Ser Lys Ala Cys Ile Pro Tyr Leu
 115 120 125

Lys Lys Ser Lys Val Ala His Ile Leu Asn Ile Ser Pro Pro Leu Asn
 130 135 140

Leu Asn Pro Val Trp Phe Lys Gln His Cys Ala Tyr Thr Ile Ala Lys
 145 150 155 160

Tyr Gly Met Ser Met Tyr Val Leu Gly Met Ala Glu Glu Phe Lys Gly
 165 170 175
 Glu Ile Ala Val Asn Ala Leu Trp Pro Lys Thr Ala Ile His Thr Ala
 180 185 190
 Ala Met Asp Met Leu Gly Gly Pro Gly Ile Glu Ser Gln Cys Arg Lys
 195 200 205
 Val Asp Ile Ile Ala Asp Ala Ala Tyr Ser Ile Phe Gln Lys Pro Lys
 210 215 220
 Ser Phe Thr Gly Asn Phe Val Ile Asp Glu Asn Ile Leu Lys Glu Glu
 225 230 235 240
 Gly Ile Glu Asn Phe Asp Val Tyr Ala Ile Lys Pro Gly His Pro Leu
 245 250 255
 Gln Pro Asp Phe Phe Leu Asp Glu Tyr Pro Glu Ala Val Ser Lys Lys
 260 265 270
 Val Glu Ser Thr Gly Ala Val Pro Glu Phe Lys Glu Glu Lys Leu Gln
 275 280 285
 Leu Gln Pro Lys Pro Arg Ser Gly Ala Val Glu Glu Thr Phe Arg Ile
 290 295 300
 Val Lys Asp Ser Leu Ser Asp Asp Val Val Lys Ala Thr Gln Ala Ile
 305 310 315 320
 Tyr Leu Phe Glu Leu Ser Gly Glu Asp Gly Gly Thr Trp Phe Leu Asp
 325 330 335
 Leu Lys Ser Lys Gly Gly Asn Val Gly Tyr Gly Glu Pro Ser Asp Gln
 340 345 350
 Ala Asp Val Val Met Ser Met Thr Thr Asp Asp Phe Val Lys Met Phe
 355 360 365
 Ser Gly Lys Leu Lys Pro Thr Met Ala Phe Met Ser Gly Lys Leu Lys
 370 375 380
 Ile Lys Gly Asn Met Ala Leu Ala Ile Lys Leu Glu Lys Leu Met Asn
 385 390 395 400
 Gln Met Asn Ala Arg Leu
 405

<210> 140
 <211> 422
 <212> PRT
 <213> Homo sapiens

<400> 140

Ser Thr Lys Val Met Leu Pro Asn Thr Gly Arg Leu Ala Gly Cys Thr
 1 5 10 15
 Val Phe Ile Thr Gly Ala Ser Arg Gly Ile Gly Lys Ala Ile Ala Leu
 20 25 30
 Lys Ala Ala Lys Asp Gly Ala Asn Ile Val Ile Ala Ala Lys Thr Ala
 35 40 45
 Gln Pro His Pro Lys Leu Leu Gly Thr Ile Tyr Thr Ala Ala Glu Glu
 50 55 60
 Ile Glu Ala Val Gly Gly Lys Ala Leu Pro Cys Ile Val Asp Val Arg
 65 70 75 80
 Asp Glu Gln Gln Ile Ser Ala Ala Val Glu Lys Ala Ile Lys Lys Phe
 85 90 95
 Gly Gly Ile Asp Ile Leu Val Asn Asn Ala Ser Ala Ile Ser Leu Thr
 100 105 110
 Asn Thr Leu Asp Thr Pro Thr Lys Arg Leu Asp Leu Met Met Asn Val
 115 120 125
 Asn Thr Arg Gly Thr Tyr Leu Ala Ser Lys Ala Cys Ile Pro Tyr Leu
 130 135 140
 Lys Lys Ser Lys Val Ala His Ile Leu Asn Ile Ser Pro Pro Leu Asn
 145 150 155 160
 Leu Asn Pro Val Trp Phe Lys Gln His Cys Ala Tyr Thr Ile Ala Lys
 165 170 175
 Tyr Gly Met Ser Met Tyr Val Leu Gly Met Ala Glu Glu Phe Lys Gly
 180 185 190
 Glu Ile Ala Val Asn Ala Leu Trp Pro Lys Thr Ala Ile His Thr Ala
 195 200 205
 Ala Met Asp Met Leu Gly Gly Pro Gly Ile Glu Ser Gln Cys Arg Lys

210	215	220
Val Asp Ile Ile Ala Asp Ala Ala Tyr Ser Ile Phe Gln Lys Pro Lys		
225	230	235 240
Ser Phe Thr Gly Asn Phe Val Ile Asp Glu Asn Ile Leu Lys Glu Glu		
	245	250 255
Gly Ile Glu Asn Phe Asp Val Tyr Ala Ile Lys Pro Gly His Pro Leu		
	260	265 270
Gln Pro Asp Phe Phe Leu Asp Glu Tyr Pro Glu Ala Val Ser Lys Lys		
	275	280 285
Val Glu Ser Thr Gly Ala Val Pro Glu Phe Lys Glu Glu Lys Leu Gln		
	290	295 300
Leu Gln Pro Lys Pro Arg Ser Gly Ala Val Glu Glu Thr Phe Arg Ile		
305	310	315 320
Val Lys Asp Ser Leu Ser Asp Asp Val Val Lys Ala Thr Gln Ala Ile		
	325	330 335
Tyr Leu Phe Glu Leu Ser Gly Glu Asp Gly Gly Thr Trp Phe Leu Asp		
	340	345 350
Leu Lys Ser Lys Gly Gly Asn Val Gly Tyr Gly Glu Pro Ser Asp Gln		
	355	360 365
Ala Asp Val Val Met Ser Met Thr Thr Asp Asp Phe Val Lys Met Phe		
	370	375 380
Ser Gly Lys Leu Lys Pro Thr Met Ala Phe Met Ser Gly Lys Leu Lys		
385	390	395 400
Ile Lys Gly Asn Met Ala Leu Ala Ile Lys Leu Glu Lys Leu Met Asn		
	405	410 415
Gln Met Asn Ala Arg Leu		
	420	

<210> 141

<211> 424

<212> PRT

<213> Homo sapiens

<400> 141

Thr Glu Val Ser Val Asp Leu Pro Lys His Ala Lys Gly Gln Asp Leu
 1 5 10 15

Phe Asp Gln Ile Val Tyr His Leu Asp Leu Val Glu Thr Asp Tyr Phe
 20 25 30

Gly Leu Gln Phe Leu Asp Ser Ala Gln Val Ala His Trp Leu Asp His
 35 40 45

Ala Lys Pro Ile Lys Lys Gln Met Lys Ile Gly Pro Ala Tyr Ala Leu
 50 55 60

His Phe Arg Val Lys Tyr Tyr Ser Ser Glu Pro Asn Asn Leu Arg Glu
 65 70 75 80

Glu Phe Thr Arg Tyr Leu Phe Val Leu Gln Leu Arg His Asp Ile Leu
 85 90 95

Ser Gly Lys Leu Lys Cys Pro Tyr Glu Thr Ala Val Glu Leu Ala Ala
 100 105 110

Leu Cys Leu Gln Ala Glu Leu Gly Glu Cys Glu Leu Pro Glu His Thr
 115 120 125

Pro Glu Leu Val Ser Glu Phe Arg Phe Ile Pro Asn Gln Thr Glu Ala
 130 135 140

Met Glu Phe Asp Ile Phe Gln Arg Trp Lys Glu Cys Arg Gly Lys Ser
 145 150 155 160

Pro Ala Gln Ala Glu Leu Ser Tyr Leu Asn Lys Ala Lys Trp Leu Glu
 165 170 175

Met Tyr Gly Val Asp Met His Val Val Arg Gly Arg Asp Gly Cys Glu
 180 185 190

Tyr Ser Leu Gly Leu Thr Pro Thr Gly Ile Leu Ile Phe Glu Gly Ala
 195 200 205

Asn Lys Ile Gly Leu Phe Phe Trp Pro Lys Ile Thr Lys Met Asp Phe
 210 215 220

Lys Lys Ser Lys Leu Thr Leu Val Val Val Glu Asp Asp Asp Gln Gly
 225 230 235 240

Arg Glu Gln Glu His Thr Phe Val Phe Arg Leu Asp Ser Ala Arg Thr
 245 250 255

Cys Lys His Leu Trp Lys Cys Ala Val Glu His His Ala Phe Phe Arg
 260 265 270
 Leu Arg Thr Pro Gly Asn Ser Lys Ser Asn Arg Ser Asp Phe Ile Arg
 275 280 285
 Leu Gly Ser Arg Phe Arg Phe Ser Gly Arg Thr Glu Tyr Gln Ala Thr
 290 295 300
 His Gly Ser Arg Leu Arg Arg Thr Ser Thr Phe Glu Arg Lys Pro Ser
 305 310 315 320
 Lys Arg Tyr Pro Ser Arg Arg His Ser Thr Phe Lys Ala Ser Asn Pro
 325 330 335
 Val Ile Ala Ala Gln Leu Cys Ser Lys Thr Asn Pro Glu Val His Asn
 340 345 350
 Tyr Gln Pro Gln Tyr His Pro Asn Ile His Pro Ser Gln Pro Arg Trp
 355 360 365
 His Pro His Ser Pro Asn Val Arg Pro Ser Phe Gln Asp Asp Arg Ser
 370 375 380
 His Trp Lys Ala Ser Ala Ser Gly Asp Asp Ser His Phe Asp Tyr Val
 385 390 395 400
 His Asp Gln Asn Gln Lys Asn Leu Gly Gly Met Gln Ser Met Met Tyr
 405 410 415
 Arg Asp Lys Leu Met Thr Ala Leu
 420

<210> 142

<211> 455

<212> PRT

<213> Homo sapiens

<400> 142

Gly Pro Leu Leu Thr Gly Gly Ala Ala Val His Ile Ser Ala Ala Gly
 1 5 10 15

Ala Ala Lys Ala Thr Leu Tyr Cys Arg Val Phe Leu Leu Asp Gly Thr
 20 25 30

Glu Val Ser Val Asp Leu Pro Lys His Ala Lys Gly Gln Asp Leu Phe
 35 40 45
 Asp Gln Ile Val Tyr His Leu Asp Leu Val Glu Thr Asp Tyr Phe Gly
 50 55 60
 Leu Gln Phe Leu Asp Ser Ala Gln Val Ala His Trp Leu Asp His Ala
 65 70 75 80
 Lys Pro Ile Lys Lys Gln Met Lys Ile Gly Pro Ala Tyr Ala Leu His
 85 90 95
 Phe Arg Val Lys Tyr Tyr Ser Ser Glu Pro Asn Asn Leu Arg Glu Glu
 100 105 110
 Phe Thr Arg Tyr Leu Phe Val Leu Gln Leu Arg His Asp Ile Leu Ser
 115 120 125
 Gly Lys Leu Lys Cys Pro Tyr Glu Thr Ala Val Glu Leu Ala Ala Leu
 130 135 140
 Cys Leu Gln Ala Glu Leu Gly Glu Cys Glu Leu Pro Glu His Thr Pro
 145 150 155 160
 Glu Leu Val Ser Glu Phe Arg Phe Ile Pro Asn Gln Thr Glu Ala Met
 165 170 175
 Glu Phe Asp Ile Phe Gln Arg Trp Lys Glu Cys Arg Gly Lys Ser Pro
 180 185 190
 Ala Gln Ala Glu Leu Ser Tyr Leu Asn Lys Ala Lys Trp Leu Glu Met
 195 200 205
 Tyr Gly Val Asp Met His Val Val Arg Gly Arg Asp Gly Cys Glu Tyr
 210 215 220
 Ser Leu Gly Leu Thr Pro Thr Gly Ile Leu Ile Phe Glu Gly Ala Asn
 225 230 235 240
 Lys Ile Gly Leu Phe Phe Trp Pro Lys Ile Thr Lys Met Asp Phe Lys
 245 250 255
 Lys Ser Lys Leu Thr Leu Val Val Val Glu Asp Asp Asp Gln Gly Arg
 260 265 270
 Glu Gln Glu His Thr Phe Val Phe Arg Leu Asp Ser Ala Arg Thr Cys
 275 280 285

Lys His Leu Trp Lys Cys Ala Val Glu His His Ala Phe Phe Arg Leu
 290 295 300
 Arg Thr Pro Gly Asn Ser Lys Ser Asn Arg Ser Asp Phe Ile Arg Leu
 305 310 315 320
 Gly Ser Arg Phe Arg Phe Ser Gly Arg Thr Glu Tyr Gln Ala Thr His
 325 330 335
 Gly Ser Arg Leu Arg Arg Thr Ser Thr Phe Glu Arg Lys Pro Ser Lys
 340 345 350
 Arg Tyr Pro Ser Arg Arg His Ser Thr Phe Lys Ala Ser Asn Pro Val
 355 360 365
 Ile Ala Ala Gln Leu Cys Ser Lys Thr Asn Pro Glu Val His Asn Tyr
 370 375 380
 Gln Pro Gln Tyr His Pro Asn Ile His Pro Ser Gln Pro Arg Trp His
 385 390 395 400
 Pro His Ser Pro Asn Val Arg Pro Ser Phe Gln Asp Asp Arg Ser His
 405 410 415
 Trp Lys Ala Ser Ala Ser Gly Asp Asp Ser His Phe Asp Tyr Val His
 420 425 430
 Asp Gln Asn Gln Lys Asn Leu Gly Gly Met Gln Ser Met Met Tyr Arg
 435 440 445
 Asp Lys Leu Met Thr Ala Leu
 450 455

<210> 143
 <211> 232
 <212> PRT
 <213> Homo sapiens

<400> 143
 Glu Asp Asp Tyr Gly Asp Asp Pro Ser Thr Asn Ser Phe Glu Lys Glu
 1 5 10 15
 Lys Gln Asp Tyr Val Tyr Cys Leu Glu Ser Ser Leu Gln Thr Tyr Asn
 20 25 30
 Pro Asp Tyr Val Leu Met Val Glu Asp Asp Ala Val Pro Glu Glu Gln

35 40 45
 Ile Phe Pro Val Leu Glu His Leu Leu Arg Ala Arg Phe Ser Glu Pro
 50 55 60
 His Leu Arg Asp Ala Leu Tyr Leu Lys Leu Tyr His Pro Glu Arg Leu
 65 70 75 80
 Gln His Tyr Ile Asn Pro Glu Pro Met Arg Ile Leu Glu Trp Val Gly
 85 90 95
 Val Gly Met Leu Leu Gly Pro Leu Leu Thr Trp Ile Tyr Met Arg Phe
 100 105 110
 Ala Ser Arg Pro Gly Phe Ser Trp Pro Val Met Leu Phe Phe Ser Leu
 115 120 125
 Tyr Ser Met Gly Leu Val Glu Leu Val Gly Arg His Tyr Phe Leu Glu
 130 135 140
 Leu Arg Arg Leu Ser Pro Ser Leu Tyr Ser Val Val Pro Ala Ser Gln
 145 150 155 160
 Cys Cys Thr Pro Ala Met Leu Phe Pro Ala Pro Ala Ala Arg Arg Thr
 165 170 175
 Leu Thr Tyr Leu Ser Gln Val Tyr Cys His Lys Gly Phe Gly Lys Asp
 180 185 190
 Met Ala Leu Tyr Ser Leu Leu Arg Ala Lys Gly Glu Arg Ala Tyr Val
 195 200 205
 Val Glu Pro Asn Leu Val Lys His Ile Gly Leu Phe Ser Ser Leu Arg
 210 215 220
 Tyr Asn Phe His Pro Ser Leu Leu
 225 230

<210> 144

<211> 263

<212> PRT

<213> Homo sapiens

<400> 144

Leu Phe Leu Cys Asn Val Glu Arg Ser Val Ser His Phe Asp Ala Lys
 1 5 10 15

Leu Leu Ser Lys Tyr Val Pro Val Ala Asn Arg Tyr Glu Gly Thr Glu
 20 25 30

Asp Asp Tyr Gly Asp Asp Pro Ser Thr Asn Ser Phe Glu Lys Glu Lys
 35 40 45

Gln Asp Tyr Val Tyr Cys Leu Glu Ser Ser Leu Gln Thr Tyr Asn Pro
 50 55 60

Asp Tyr Val Leu Met Val Glu Asp Asp Ala Val Pro Glu Glu Gln Ile
 65 70 75 80

Phe Pro Val Leu Glu His Leu Leu Arg Ala Arg Phe Ser Glu Pro His
 85 90 95

Leu Arg Asp Ala Leu Tyr Leu Lys Leu Tyr His Pro Glu Arg Leu Gln
 100 105 110

His Tyr Ile Asn Pro Glu Pro Met Arg Ile Leu Glu Trp Val Gly Val
 115 120 125

Gly Met Leu Leu Gly Pro Leu Leu Thr Trp Ile Tyr Met Arg Phe Ala
 130 135 140

Ser Arg Pro Gly Phe Ser Trp Pro Val Met Leu Phe Phe Ser Leu Tyr
 145 150 155 160

Ser Met Gly Leu Val Glu Leu Val Gly Arg His Tyr Phe Leu Glu Leu
 165 170 175

Arg Arg Leu Ser Pro Ser Leu Tyr Ser Val Val Pro Ala Ser Gln Cys
 180 185 190

Cys Thr Pro Ala Met Leu Phe Pro Ala Pro Ala Ala Arg Arg Thr Leu
 195 200 205

Thr Tyr Leu Ser Gln Val Tyr Cys His Lys Gly Phe Gly Lys Asp Met
 210 215 220

Ala Leu Tyr Ser Leu Leu Arg Ala Lys Gly Glu Arg Ala Tyr Val Val
 225 230 235 240

Glu Pro Asn Leu Val Lys His Ile Gly Leu Phe Ser Ser Leu Arg Tyr
 245 250 255

Asn Phe His Pro Ser Leu Leu
 260

<210> 145
 <211> 68
 <212> PRT
 <213> Homo sapiens

<400> 145
 Leu Tyr Ser Lys Leu Ala Leu Gln Gly Val Ser Ser Val Tyr Pro Ala
 1 5 10 15
 Pro Val Cys Leu Ala Thr Pro Phe Leu Gly Arg Val Asp Leu Ala Val
 20 25 30
 His Val Ile Asp Gly His Ser Val Tyr Leu Asp Ser Phe Leu Phe Leu
 35 40 45
 Cys Gln Gln Val Ile Gly Met Phe Gly Phe Lys Asn Met Ser Val Leu
 50 55 60
 Cys Asn Phe Met
 65

<210> 146
 <211> 720
 <212> PRT
 <213> Homo sapiens

<400> 146
 Thr Cys His Glu His Ala Thr Cys Gln Gln Arg Glu Gly Lys Lys Ile
 1 5 10 15
 Cys Ile Cys Asn Tyr Gly Phe Val Gly Asn Gly Arg Thr Gln Cys Val
 20 25 30
 Asp Lys Asn Glu Cys Gln Phe Gly Ala Thr Leu Val Cys Gly Asn His
 35 40 45
 Thr Ser Cys His Asn Thr Pro Gly Gly Phe Tyr Cys Ile Cys Leu Glu
 50 55 60
 Gly Tyr Arg Ala Thr Asn Asn Asn Lys Thr Phe Ile Pro Asn Asp Gly
 65 70 75 80
 Thr Phe Cys Thr Asp Ile Asp Glu Cys Glu Val Ser Gly Leu Cys Arg

85	90	95
His Gly Gly Arg Cys Val Asn Thr	His Gly Ser Phe Glu Cys Tyr Cys	
100	105	110
Met Asp Gly Tyr Leu Pro Arg Asn Gly Pro Glu Pro Phe His Pro Thr		
115	120	125
Thr Asp Ala Thr Ser Cys Thr Glu Ile Asp Cys Gly Thr Pro Pro Glu		
130	135	140
Val Pro Asp Gly Tyr Ile Ile Gly Asn Tyr Thr Ser Ser Leu Gly Ser		
145	150	155
Gln Val Arg Tyr Ala Cys Arg Glu Gly Phe Phe Ser Val Pro Glu Asp		
165	170	175
Thr Val Ser Ser Cys Thr Gly Leu Gly Thr Trp Glu Ser Pro Lys Leu		
180	185	190
His Cys Gln Glu Ile Asn Cys Gly Asn Pro Pro Glu Met Arg His Ala		
195	200	205
Ile Leu Val Gly Asn His Ser Ser Arg Leu Gly Gly Val Ala Arg Tyr		
210	215	220
Val Cys Gln Glu Gly Phe Glu Ser Pro Gly Gly Lys Ile Thr Ser Val		
225	230	235
Cys Thr Glu Lys Gly Thr Trp Arg Glu Ser Thr Leu Thr Cys Thr Glu		
245	250	255
Ile Leu Thr Lys Ile Asn Asp Val Ser Leu Phe Asn Asp Thr Cys Val		
260	265	270
Arg Trp Gln Ile Asn Ser Arg Arg Ile Asn Pro Lys Ile Ser Tyr Val		
275	280	285
Ile Ser Ile Lys Gly Gln Arg Leu Asp Pro Met Glu Ser Val Arg Glu		
290	295	300
Glu Thr Val Asn Leu Thr Thr Asp Ser Arg Thr Pro Glu Val Cys Leu		
305	310	315
Ala Leu Tyr Pro Gly Thr Asn Tyr Thr Val Asn Ile Ser Thr Ala Pro		
325	330	335
Pro Arg Arg Ser Met Pro Ala Val Ile Gly Phe Gln Thr Ala Gly Ser		

340	345	350
Arg Asn Ser His Gly Arg Asn Thr His Ile Lys Glu Val Asp Leu Leu		
355	360	365
Glu Asp Asp Gly Ser Phe Asn Ile Ser Ile Phe Asn Glu Thr Cys Leu		
370	375	380
Lys Leu Asn Arg Arg Ser Arg Lys Val Gly Ser Glu His Met Tyr Gln		
385	390	395
Phe Thr Val Leu Gly Gln Arg Trp Tyr Leu Ala Asn Phe Ser His Ala		
405	410	415
Thr Ser Phe Asn Phe Thr Thr Arg Glu Gln Val Pro Val Val Cys Leu		
420	425	430
Asp Leu Tyr Pro Thr Thr Asp Tyr Thr Val Asn Val Thr Leu Leu Arg		
435	440	445
Ser Pro Lys Arg His Ser Val Gln Ile Thr Ile Ala Thr Pro Pro Ala		
450	455	460
Val Lys Gln Thr Ile Ser Asn Ile Ser Gly Phe Asn Glu Thr Cys Leu		
465	470	475
Arg Trp Arg Ser Ile Lys Thr Ala Asp Met Glu Glu Met Tyr Leu Phe		
485	490	495
His Ile Trp Gly Gln Arg Trp Tyr Gln Lys Glu Phe Ala Gln Glu Met		
500	505	510
Thr Phe Asn Ile Ser Ser Ser Ser Arg Asp Pro Glu Val Cys Leu Asp		
515	520	525
Leu Arg Pro Gly Thr Asn Tyr Asn Val Ser Leu Arg Ala Leu Ser Ser		
530	535	540
Glu Leu Pro Val Val Ile Ser Leu Thr Thr Gln Ile Thr Glu Pro Pro		
545	550	555
Leu Pro Glu Val Glu Phe Phe Thr Val His Arg Gly Pro Leu Pro Arg		
565	570	575
Leu Arg Leu Arg Lys Ala Lys Glu Lys Asn Gly Pro Ile Ser Ser Tyr		
580	585	590
Gln Val Leu Val Leu Pro Leu Ala Leu Gln Ser Thr Phe Ser Cys Asp		

595 600 605
 Ser Glu Gly Ala Ser Ser Phe Phe Ser Asn Ala Ser Asp Ala Asp Gly
 610 615 620
 Tyr Val Ala Ala Glu Leu Leu Ala Lys Asp Val Pro Asp Asp Ala Met
 625 630 635 640
 Glu Ile Pro Ile Gly Asp Arg Leu Tyr Tyr Gly Glu Tyr Tyr Asn Ala
 645 650 655
 Pro Leu Lys Arg Gly Ser Asp Tyr Cys Ile Ile Leu Arg Ile Thr Ser
 660 665 670
 Glu Trp Asn Lys Val Arg Arg His Ser Cys Ala Val Trp Ala Gln Val
 675 680 685
 Lys Asp Ser Ser Leu Met Leu Leu Gln Met Ala Gly Val Gly Leu Gly
 690 695 700
 Ser Leu Ala Val Val Ile Ile Leu Thr Phe Leu Ser Phe Ser Ala Val
 705 710 715 720

<210> 147
 <211> 707
 <212> PRT
 <213> Homo sapiens

<400> 147
 Thr Cys His Glu His Ala Thr Cys Gln Gln Arg Glu Gly Lys Lys Ile
 1 5 10 15
 Cys Ile Cys Asn Tyr Gly Phe Val Gly Asn Gly Arg Thr Gln Cys Val
 20 25 30
 Asp Lys Asn Glu Cys Gln Phe Gly Ala Thr Leu Val Cys Gly Asn His
 35 40 45
 Thr Ser Cys His Asn Thr Pro Gly Gly Phe Tyr Cys Ile Cys Leu Glu
 50 55 60
 Gly Tyr Arg Ala Thr Asn Asn Asn Lys Thr Phe Ile Pro Asn Asp Gly
 65 70 75 80

Thr	Phe	Cys	Thr	Asp	Ile	Asp	Glu	Cys	Glu	Val	Ser	Gly	Leu	Cys	Arg	85	90	95
His	Gly	Gly	Arg	Cys	Val	Asn	Thr	His	Gly	Ser	Phe	Glu	Cys	Tyr	Cys	100	105	110
Met	Asp	Gly	Tyr	Leu	Pro	Arg	Asn	Gly	Pro	Glu	Pro	Phe	His	Pro	Thr	115	120	125
Thr	Asp	Ala	Thr	Ser	Cys	Thr	Glu	Ile	Asp	Cys	Gly	Thr	Pro	Pro	Glu	130	135	140
Val	Pro	Asp	Gly	Tyr	Ile	Ile	Gly	Asn	Tyr	Thr	Ser	Ser	Leu	Gly	Ser	145	150	155
Gln	Val	Arg	Tyr	Ala	Cys	Arg	Glu	Gly	Phe	Phe	Ser	Val	Pro	Glu	Asp	165	170	175
Thr	Val	Ser	Ser	Cys	Thr	Gly	Leu	Gly	Thr	Trp	Glu	Ser	Pro	Lys	Leu	180	185	190
His	Cys	Gln	Glu	Ile	Asn	Cys	Gly	Asn	Pro	Pro	Glu	Met	Arg	His	Ala	195	200	205
Ile	Leu	Val	Gly	Asn	His	Ser	Ser	Arg	Leu	Gly	Gly	Val	Ala	Arg	Tyr	210	215	220
Val	Cys	Gln	Glu	Gly	Phe	Glu	Ser	Pro	Gly	Gly	Lys	Ile	Thr	Ser	Val	225	230	235
Cys	Thr	Glu	Lys	Gly	Thr	Trp	Arg	Glu	Ser	Thr	Leu	Thr	Cys	Thr	Glu	245	250	255
Ile	Leu	Thr	Lys	Ile	Asn	Asp	Val	Ser	Leu	Phe	Asn	Asp	Thr	Cys	Val	260	265	270
Arg	Trp	Gln	Ile	Asn	Ser	Arg	Arg	Ile	Asn	Pro	Lys	Ile	Ser	Tyr	Val	275	280	285
Ile	Ser	Ile	Lys	Gly	Gln	Arg	Leu	Asp	Pro	Met	Glu	Ser	Val	Arg	Glu	290	295	300
Glu	Thr	Val	Asn	Leu	Thr	Thr	Asp	Ser	Arg	Thr	Pro	Glu	Val	Cys	Leu	305	310	315
Ala	Leu	Tyr	Pro	Gly	Thr	Asn	Tyr	Thr	Val	Asn	Ile	Ser	Thr	Ala	Pro	325	330	335

Pro	Arg	Arg	Ser	Met	Pro	Ala	Val	Ile	Gly	Phe	Gln	Thr	Ala	Glu	Val
				340					345					350	
Asp	Leu	Leu	Glu	Asp	Asp	Gly	Ser	Phe	Asn	Ile	Ser	Ile	Phe	Asn	Glu
		355					360					365			
Thr	Cys	Leu	Lys	Leu	Asn	Arg	Arg	Ser	Arg	Lys	Val	Gly	Ser	Glu	His
	370					375					380				
Met	Tyr	Gln	Phe	Thr	Val	Leu	Gly	Gln	Arg	Trp	Tyr	Leu	Ala	Asn	Phe
385					390					395					400
Ser	His	Ala	Thr	Ser	Phe	Asn	Phe	Thr	Thr	Arg	Glu	Gln	Val	Pro	Val
				405					410					415	
Val	Cys	Leu	Asp	Leu	Tyr	Pro	Thr	Thr	Asp	Tyr	Thr	Val	Asn	Val	Thr
			420					425					430		
Leu	Leu	Arg	Ser	Pro	Lys	Arg	His	Ser	Val	Gln	Ile	Thr	Ile	Ala	Thr
		435					440					445			
Pro	Pro	Ala	Val	Lys	Gln	Thr	Ile	Ser	Asn	Ile	Ser	Gly	Phe	Asn	Glu
	450					455					460				
Thr	Cys	Leu	Arg	Trp	Arg	Ser	Ile	Lys	Thr	Ala	Asp	Met	Glu	Glu	Met
465					470					475					480
Tyr	Leu	Phe	His	Ile	Trp	Gly	Gln	Arg	Trp	Tyr	Gln	Lys	Glu	Phe	Ala
			485						490					495	
Gln	Glu	Met	Thr	Phe	Asn	Ile	Ser	Ser	Ser	Ser	Arg	Asp	Pro	Glu	Val
			500					505					510		
Cys	Leu	Asp	Leu	Arg	Pro	Gly	Thr	Asn	Tyr	Asn	Val	Ser	Leu	Arg	Ala
		515					520					525			
Leu	Ser	Ser	Glu	Leu	Pro	Val	Val	Ile	Ser	Leu	Thr	Thr	Gln	Ile	Thr
	530					535					540				
Glu	Pro	Pro	Leu	Pro	Glu	Val	Glu	Phe	Phe	Thr	Val	His	Arg	Gly	Pro
545					550					555					560
Leu	Pro	Arg	Leu	Arg	Leu	Arg	Lys	Ala	Lys	Glu	Lys	Asn	Gly	Pro	Ile
				565					570					575	
Ser	Ser	Tyr	Gln	Val	Leu	Val	Leu	Pro	Leu	Ala	Leu	Gln	Ser	Thr	Phe
			580					585					590		

Ser Cys Asp Ser Glu Gly Ala Ser Ser Phe Phe Ser Asn Ala Ser Asp
595 600 605

Ala Asp Gly Tyr Val Ala Ala Glu Leu Leu Ala Lys Asp Val Pro Asp
610 615 620

Asp Ala Met Glu Ile Pro Ile Gly Asp Arg Leu Tyr Tyr Gly Glu Tyr
625 630 635 640

Tyr Asn Ala Pro Leu Lys Arg Gly Ser Asp Tyr Cys Ile Ile Leu Arg
645 650 655

Ile Thr Ser Glu Trp Asn Lys Val Arg Arg His Ser Cys Ala Val Trp
660 665 670

Ala Gln Val Lys Asp Ser Ser Leu Met Leu Leu Gln Met Ala Gly Val
675 680 685

Gly Leu Gly Ser Leu Ala Val Val Ile Ile Leu Thr Phe Leu Ser Phe
690 695 700

Ser Ala Val
705

<210> 148

<211> 713

<212> PRT

<213> Homo sapiens

<400> 148

Ser Leu Asp Val Cys Ala Thr Cys His Glu His Ala Thr Cys Gln Gln
1 5 10 15

Arg Glu Gly Lys Lys Ile Cys Ile Cys Asn Tyr Gly Phe Val Gly Asn
20 25 30

Gly Arg Thr Gln Cys Val Asp Lys Asn Glu Cys Gln Phe Gly Ala Thr
35 40 45

Leu Val Cys Gly Asn His Thr Ser Cys His Asn Thr Pro Gly Gly Phe
50 55 60

Tyr Cys Ile Cys Leu Glu Gly Tyr Arg Ala Thr Asn Asn Asn Lys Thr
65 70 75 80

Phe	Ile	Pro	Asn	Asp	Gly	Thr	Phe	Cys	Thr	Asp	Ile	Asp	Glu	Cys	Glu	85	90	95
Val	Ser	Gly	Leu	Cys	Arg	His	Gly	Gly	Arg	Cys	Val	Asn	Thr	His	Gly	100	105	110
Ser	Phe	Glu	Cys	Tyr	Cys	Met	Asp	Gly	Tyr	Leu	Pro	Arg	Asn	Gly	Pro	115	120	125
Glu	Pro	Phe	His	Pro	Thr	Thr	Asp	Ala	Thr	Ser	Cys	Thr	Glu	Ile	Asp	130	135	140
Cys	Gly	Thr	Pro	Pro	Glu	Val	Pro	Asp	Gly	Tyr	Ile	Ile	Gly	Asn	Tyr	145	150	155
Thr	Ser	Ser	Leu	Gly	Ser	Gln	Val	Arg	Tyr	Ala	Cys	Arg	Glu	Gly	Phe	165	170	175
Phe	Ser	Val	Pro	Glu	Asp	Thr	Val	Ser	Ser	Cys	Thr	Gly	Leu	Gly	Thr	180	185	190
Trp	Glu	Ser	Pro	Lys	Leu	His	Cys	Gln	Glu	Ile	Asn	Cys	Gly	Asn	Pro	195	200	205
Pro	Glu	Met	Arg	His	Ala	Ile	Leu	Val	Gly	Asn	His	Ser	Ser	Arg	Leu	210	215	220
Gly	Gly	Val	Ala	Arg	Tyr	Val	Cys	Gln	Glu	Gly	Phe	Glu	Ser	Pro	Gly	225	230	235
Gly	Lys	Ile	Thr	Ser	Val	Cys	Thr	Glu	Lys	Gly	Thr	Trp	Arg	Glu	Ser	245	250	255
Thr	Leu	Thr	Cys	Thr	Glu	Ile	Leu	Thr	Lys	Ile	Asn	Asp	Val	Ser	Leu	260	265	270
Phe	Asn	Asp	Thr	Cys	Val	Arg	Trp	Gln	Ile	Asn	Ser	Arg	Arg	Ile	Asn	275	280	285
Pro	Lys	Ile	Ser	Tyr	Val	Ile	Ser	Ile	Lys	Gly	Gln	Arg	Leu	Asp	Pro	290	295	300
Met	Glu	Ser	Val	Arg	Glu	Glu	Thr	Val	Asn	Leu	Thr	Thr	Asp	Ser	Arg	305	310	315
Thr	Pro	Glu	Val	Cys	Leu	Ala	Leu	Tyr	Pro	Gly	Thr	Asn	Tyr	Thr	Val	325	330	335

Asn Ile Ser Thr Ala Pro Pro Arg Arg Ser Met Pro Ala Val Ile Gly
 340 345 350
 Phe Gln Thr Ala Glu Val Asp Leu Leu Glu Asp Asp Gly Ser Phe Asn
 355 360 365
 Ile Ser Ile Phe Asn Glu Thr Cys Leu Lys Leu Asn Arg Arg Ser Arg
 370 375 380
 Lys Val Gly Ser Glu His Met Tyr Gln Phe Thr Val Leu Gly Gln Arg
 385 390 395 400
 Trp Tyr Leu Ala Asn Phe Ser His Ala Thr Ser Phe Asn Phe Thr Thr
 405 410 415
 Arg Glu Gln Val Pro Val Val Cys Leu Asp Leu Tyr Pro Thr Thr Asp
 420 425 430
 Tyr Thr Val Asn Val Thr Leu Leu Arg Ser Pro Lys Arg His Ser Val
 435 440 445
 Gln Ile Thr Ile Ala Thr Pro Pro Ala Val Lys Gln Thr Ile Ser Asn
 450 455 460
 Ile Ser Gly Phe Asn Glu Thr Cys Leu Arg Trp Arg Ser Ile Lys Thr
 465 470 475 480
 Ala Asp Met Glu Glu Met Tyr Leu Phe His Ile Trp Gly Gln Arg Trp
 485 490 495
 Tyr Gln Lys Glu Phe Ala Gln Glu Met Thr Phe Asn Ile Ser Ser Ser
 500 505 510
 Ser Arg Asp Pro Glu Val Cys Leu Asp Leu Arg Pro Gly Thr Asn Tyr
 515 520 525
 Asn Val Ser Leu Arg Ala Leu Ser Ser Glu Leu Pro Val Val Ile Ser
 530 535 540
 Leu Thr Thr Gln Ile Thr Glu Pro Pro Leu Pro Glu Val Glu Phe Phe
 545 550 555 560
 Thr Val His Arg Gly Pro Leu Pro Arg Leu Arg Leu Arg Lys Ala Lys
 565 570 575
 Glu Lys Asn Gly Pro Ile Ser Ser Tyr Gln Val Leu Val Leu Pro Leu
 580 585 590

Ala Leu Gln Ser Thr Phe Ser Cys Asp Ser Glu Gly Ala Ser Ser Phe
 595 600 605

Phe Ser Asn Ala Ser Asp Ala Asp Gly Tyr Val Ala Ala Glu Leu Leu
 610 615 620

Ala Lys Asp Val Pro Asp Asp Ala Met Glu Ile Pro Ile Gly Asp Arg
 625 630 635 640

Leu Tyr Tyr Gly Glu Tyr Tyr Asn Ala Pro Leu Lys Arg Gly Ser Asp
 645 650 655

Tyr Cys Ile Ile Leu Arg Ile Thr Ser Glu Trp Asn Lys Val Arg Arg
 660 665 670

His Ser Cys Ala Val Trp Ala Gln Val Lys Asp Ser Ser Leu Met Leu
 675 680 685

Leu Gln Met Ala Gly Val Gly Leu Gly Ser Leu Ala Val Val Ile Ile
 690 695 700

Leu Thr Phe Leu Ser Phe Ser Ala Val
 705 710

<210> 149
 <211> 170
 <212> PRT
 <213> Homo sapiens

<400> 149
 Arg Ser Met Thr Val Gly Gly Ala Phe His Thr Thr Glu Ala Glu Ala
 1 5 10 15

Ser Ser Gln Ser Leu Thr Gln Ile Tyr Ala Leu Pro Glu Ile Pro Gln
 20 25 30

Asp Gln Asn Ala Ala Glu Ser Trp Glu Thr Leu Glu Ala Asp Leu Ile
 35 40 45

Glu Leu Ser Gln Leu Val Thr Asp Phe Ser Leu Leu Val Asn Ser Gln
 50 55 60

Gln Glu Lys Ile Asp Ser Ile Ala Asp His Val Asn Ser Ala Ala Val
 65 70 75 80

Asn Val Glu Glu Gly Thr Lys Asn Leu Gly Lys Ala Ala Lys Tyr Lys

85 90 95
 Leu Ala Ala Leu Pro Val Ala Gly Ala Leu Ile Gly Gly Met Val Gly
 100 105 110
 Gly Pro Ile Gly Leu Leu Ala Gly Phe Lys Val Ala Gly Ile Ala Ala
 115 120 125
 Ala Leu Gly Gly Gly Val Leu Gly Phe Thr Gly Gly Lys Leu Ile Gln
 130 135 140
 Arg Lys Lys Gln Lys Met Met Glu Lys Leu Thr Ser Ser Cys Pro Asp
 145 150 155 160
 Leu Pro Ser Gln Thr Asp Lys Lys Cys Ser
 165 170

<210> 150
 <211> 318
 <212> PRT
 <213> Homo sapiens

<400> 150
 Val Glu Lys Thr Ala Val Thr Arg Glu Val Ile Gln His Phe Phe Arg
 1 5 10 15
 Met Ser Glu Asp Glu Glu Lys Val Lys Leu Arg Arg Leu Glu Pro Ala
 20 25 30
 Ile Gln Lys Phe Ile Lys Ile Val Ile Pro Thr Asp Leu Glu Arg Leu
 35 40 45
 Arg Lys His Gln Ile Asn Ile Glu Lys Tyr Gln Arg Cys Arg Ile Trp
 50 55 60
 Asp Lys Leu His Glu Glu His Ile Asn Ala Gly Arg Thr Val Gln Gln
 65 70 75 80
 Leu Arg Ser Asn Ile Arg Glu Ile Glu Lys Leu Cys Leu Lys Val Arg
 85 90 95
 Lys Asp Asp Leu Val Leu Leu Lys Arg Met Ile Asp Pro Val Lys Glu
 100 105 110
 Glu Ala Ser Ala Ala Thr Ala Glu Phe Leu Gln Leu His Leu Glu Ser
 115 120 125

Val Glu Glu Leu Lys Lys Gln Phe Asn Asp Glu Glu Thr Leu Leu Gln
 130 135 140
 Pro Pro Leu Thr Arg Ser Met Thr Val Gly Gly Ala Phe His Thr Thr
 145 150 155 160
 Glu Ala Glu Ala Ser Ser Gln Ser Leu Thr Gln Ile Tyr Ala Leu Pro
 165 170 175
 Glu Ile Pro Gln Asp Gln Asn Ala Ala Glu Ser Trp Glu Thr Leu Glu
 180 185 190
 Ala Asp Leu Ile Glu Leu Ser Gln Leu Val Thr Asp Phe Ser Leu Leu
 195 200 205
 Val Asn Ser Gln Gln Glu Lys Ile Asp Ser Ile Ala Asp His Val Asn
 210 215 220
 Ser Ala Ala Val Asn Val Glu Glu Gly Thr Lys Asn Leu Gly Lys Ala
 225 230 235 240
 Ala Lys Tyr Lys Leu Ala Ala Leu Pro Val Ala Gly Ala Leu Ile Gly
 245 250 255
 Gly Met Val Gly Gly Pro Ile Gly Leu Leu Ala Gly Phe Lys Val Ala
 260 265 270
 Gly Ile Ala Ala Ala Leu Gly Gly Gly Val Leu Gly Phe Thr Gly Gly
 275 280 285
 Lys Leu Ile Gln Arg Lys Lys Gln Lys Met Met Glu Lys Leu Thr Ser
 290 295 300
 Ser Cys Pro Asp Leu Pro Ser Gln Thr Asp Lys Lys Cys Ser
 305 310 315

<210> 151

<211> 123

<212> PRT

<213> Homo sapiens

<400> 151

His Ser Arg Cys Ser Pro Pro Leu Ser Leu Pro Met Lys Glu Glu Thr
 1 5 10 15

Thr Gly Val Cys Met Tyr Pro Pro Ile Lys Thr Arg Leu Ile Lys Thr
 20 25 30
 Phe Pro Val Asp Thr Met Thr Pro Phe Pro Asp Thr Phe Thr Thr Gly
 35 40 45
 Pro Gln Phe Thr Ala Asp Phe Arg Asp Ser Lys Leu Gln Cys Cys Pro
 50 55 60
 Gly Pro Thr Ser Pro Leu Ile Pro Ala Ala Thr Leu Arg Pro Leu Thr
 65 70 75 80
 Glu Thr Val Ser Thr Val Gln Thr Ile Tyr Thr Thr Arg Lys Pro Val
 85 90 95
 Ser Leu Ala Ala Ser Ala Glu Thr Leu Arg Gln Glu Leu Glu Arg Glu
 100 105 110
 Lys Met Met Lys Arg Leu Leu Met Thr Glu Leu
 115 120

<210> 152

<211> 58

<212> PRT

<213> Homo sapiens

<400> 152

Gln Gly Arg Ser Arg Met His Val Ser Leu Pro Thr His Leu Cys His
 1 5 10 15
 Val Gly Ser Phe Arg Ser Ala Met Val Arg Lys Glu Gln Lys Leu Leu
 20 25 30
 Val Val Phe Val Leu Phe Trp Arg Ile Cys Leu Pro Val Asn Thr Ser
 35 40 45
 Leu Pro Val Pro Gln Met Leu Pro Ala Pro
 50 55

<210> 153

<211> 36

<212> PRT

<213> Homo sapiens

<400> 153

Arg Ile Thr Thr Gly Gly Gly Arg Trp Val Pro Ala Gly Gly Gly Val
1 5 10 15

Ser Ala Pro Trp Gly Ala Arg Gln Ser Gly Gly Gly Ser Ala Ala Gly
20 25 30

Ala Gly Arg Leu
35

<210> 154

<211> 651

<212> PRT

<213> Homo sapiens

<400> 154

Pro Xaa Pro Gly Ser Phe Gly Xaa Ala Met Gly Cys Cys Ser Ser Ala
1 5 10 15

Ser Ser Ala Ala Gln Ser Ser Lys Arg Glu Trp Lys Pro Leu Glu Asp
20 25 30

Arg Ser Cys Thr Asp Ile Pro Trp Leu Leu Leu Phe Ile Leu Phe Cys
35 40 45

Ile Gly Met Gly Phe Ile Cys Gly Phe Ser Ile Ala Thr Gly Ala Ala
50 55 60

Ala Arg Leu Val Ser Gly Tyr Asp Ser Tyr Gly Asn Ile Cys Gly Gln
65 70 75 80

Lys Asn Thr Lys Leu Glu Ala Ile Pro Asn Ser Gly Met Asp His Thr
85 90 95

Gln Arg Lys Tyr Val Phe Phe Leu Asp Pro Cys Asn Leu Asp Leu Ile
100 105 110

Asn Arg Lys Ile Lys Ser Val Ala Leu Cys Val Ala Ala Cys Pro Arg
115 120 125

Gln Glu Leu Lys Thr Leu Ser Asp Val Gln Lys Phe Ala Glu Ile Asn
130 135 140

Gly Ser Ala Leu Cys Ser Tyr Xaa Leu Lys Pro Ser Glu Tyr Thr Thr
145 150 155 160

Ser Pro Lys Ser Ser Val Leu Cys Pro Lys Leu Pro Val Pro Ala Ser
 165 170 175
 Ala Pro Ile Pro Phe Phe His Arg Cys Ala Pro Val Asn Ile Ser Cys
 180 185 190
 Tyr Ala Lys Phe Ala Glu Ala Leu Ile Thr Phe Val Ser Asp Asn Ser
 195 200 205
 Val Leu His Arg Leu Ile Ser Gly Val Met Thr Ser Lys Glu Ile Ile
 210 215 220
 Leu Gly Leu Cys Leu Leu Ser Leu Val Leu Ser Met Ile Leu Met Val
 225 230 235 240
 Ile Ile Arg Tyr Ile Ser Arg Val Leu Val Trp Ile Leu Thr Ile Leu
 245 250 255
 Val Ile Leu Gly Ser Leu Gly Gly Thr Gly Val Leu Trp Trp Leu Tyr
 260 265 270
 Ala Lys Gln Arg Arg Ser Pro Lys Glu Thr Val Thr Pro Glu Gln Leu
 275 280 285
 Gln Ile Ala Glu Asp Asn Leu Arg Ala Leu Leu Ile Tyr Ala Ile Ser
 290 295 300
 Ala Thr Val Phe Thr Val Ile Leu Phe Leu Ile Met Leu Val Met Arg
 305 310 315 320
 Lys Arg Val Ala Leu Thr Ile Ala Leu Phe His Val Ala Gly Lys Val
 325 330 335
 Phe Ile His Leu Pro Leu Leu Val Phe Gln Pro Phe Trp Thr Phe Phe
 340 345 350
 Ala Leu Val Leu Phe Trp Val Tyr Trp Ile Met Thr Leu Leu Phe Leu
 355 360 365
 Gly Thr Thr Gly Ser Pro Val Gln Asn Glu Gln Gly Phe Val Glu Phe
 370 375 380
 Lys Ile Ser Gly Pro Leu Gln Tyr Met Trp Trp Tyr His Val Val Gly
 385 390 395 400
 Leu Ile Trp Ile Ser Glu Phe Ile Leu Ala Cys Gln Gln Met Thr Val
 405 410 415

Ala Gly Ala Val Val Thr Tyr Tyr Phe Thr Arg Asp Lys Arg Asn Leu
 420 425 430

Pro Phe Thr Pro Ile Leu Ala Ser Val Asn Arg Leu Ile Arg Tyr His
 435 440 445

Leu Gly Thr Val Ala Lys Gly Ser Phe Ile Ile Thr Leu Val Lys Ile
 450 455 460

Pro Arg Met Ile Leu Met Tyr Ile His Ser Gln Leu Lys Gly Lys Glu
 465 470 475 480

Asn Ala Cys Ala Arg Cys Val Leu Lys Ser Cys Ile Cys Cys Leu Trp
 485 490 495

Cys Leu Glu Lys Cys Leu Asn Tyr Leu Asn Gln Asn Ala Tyr Thr Ala
 500 505 510

Thr Ala Ile Asn Ser Thr Asn Phe Cys Thr Ser Ala Lys Asp Ala Phe
 515 520 525

Val Ile Leu Val Glu Asn Ala Leu Arg Val Ala Thr Ile Asn Thr Val
 530 535 540

Gly Asp Phe Met Leu Phe Leu Gly Lys Val Leu Ile Val Cys Ser Thr
 545 550 555 560

Gly Leu Ala Gly Ile Met Leu Leu Asn Tyr Gln Gln Asp Tyr Thr Val
 565 570 575

Trp Val Leu Pro Leu Ile Ile Val Cys Leu Phe Ala Phe Leu Val Ala
 580 585 590

His Cys Phe Leu Ser Ile Tyr Glu Met Val Val Asp Val Leu Phe Leu
 595 600 605

Cys Phe Ala Ile Asp Thr Lys Tyr Asn Asp Gly Ser Pro Gly Arg Glu
 610 615 620

Phe Tyr Met Asp Lys Val Leu Met Glu Phe Val Glu Asn Ser Arg Lys
 625 630 635 640

Ala Met Lys Glu Ala Gly Lys Gly Gly Arg Arg
 645 650

<210> 155

<211> 1066

<212> PRT

<213> Homo sapiens

<400> 155

Asn Gly Lys Phe Ser Tyr Thr Asp Leu His Tyr Gly Gln Thr Val Thr
 1 5 10 15

Tyr Ser Cys Asn Arg Gly Phe Arg Xaa Glu Gly Pro Ser Ala Leu Thr
 20 25 30

Cys Leu Glu Thr Gly Asp Trp Asp Val Asp Ala Pro Ser Cys Asn Ala
 35 40 45

Ile His Cys Asp Ser Pro Gln Pro Ile Glu Asn Gly Phe Val Glu Gly
 50 55 60

Ala Asp Tyr Ser Tyr Gly Ala Ile Ile Ile Tyr Ser Cys Phe Pro Gly
 65 70 75 80

Phe Gln Val Ala Gly His Ala Met Gln Thr Cys Glu Glu Ser Gly Trp
 85 90 95

Ser Ser Ser Ile Pro Thr Cys Met Pro Ile Asp Cys Gly Leu Pro Pro
 100 105 110

His Ile Asp Phe Gly Asp Cys Thr Lys Leu Lys Asp Asp Gln Gly Tyr
 115 120 125

Phe Glu Gln Glu Asp Asp Met Met Glu Val Pro Tyr Val Thr Pro His
 130 135 140

Pro Pro Tyr His Leu Gly Ala Val Ala Lys Thr Trp Glu Asn Thr Lys
 145 150 155 160

Glu Ser Pro Ala Thr His Ser Ser Asn Phe Leu Tyr Gly Thr Met Val
 165 170 175

Ser Tyr Thr Cys Xaa Pro Gly Tyr Glu Leu Leu Gly Asn Pro Val Leu
 180 185 190

Ile Xaa Gln Glu Asp Gly Thr Trp Asn Gly Ser Ala Pro Ser Cys Ile
 195 200 205

Ser Ile Glu Cys Asp Leu Pro Thr Ala Pro Glu Asn Gly Phe Leu Arg
 210 215 220

Phe Thr Glu Thr Ser Met Gly Ser Ala Val Gln Tyr Ser Cys Lys Pro

225		230		235		240
Gly His Ile Leu Ala Gly Ser Asp Leu Arg Leu Cys Leu Glu Asn Arg						
	245			250		255
Lys Trp Ser Gly Ala Ser Pro Arg Cys Glu Ala Ile Ser Cys Lys Lys						
	260			265		270
Pro Asn Pro Val Met Asn Gly Ser Ile Lys Gly Ser Asn Tyr Thr Tyr						
	275			280		285
Leu Ser Thr Leu Tyr Tyr Glu Cys Asp Pro Gly Tyr Val Leu Asn Gly						
	290			295		300
Thr Glu Arg Arg Thr Cys Gln Asp Asp Lys Asn Trp Asp Glu Asp Glu						
305		310		315		320
Pro Ile Cys Ile Pro Val Asp Cys Ser Ser Pro Pro Val Ser Ala Asn						
	325			330		335
Gly Gln Val Arg Gly Asp Glu Tyr Thr Phe Gln Lys Glu Ile Glu Tyr						
	340			345		350
Thr Cys Asn Glu Gly Phe Leu Leu Glu Gly Ala Arg Ser Arg Val Cys						
	355			360		365
Leu Ala Asn Gly Ser Trp Ser Gly Ala Thr Pro Asp Cys Val Pro Val						
	370			375		380
Arg Cys Ala Thr Pro Pro Gln Leu Ala Asn Gly Val Thr Glu Gly Leu						
385		390		395		400
Asp Tyr Gly Phe Met Lys Glu Val Thr Phe His Cys His Glu Gly Tyr						
	405			410		415
Ile Leu His Gly Ala Pro Lys Leu Thr Cys Gln Ser Asp Gly Asn Trp						
	420			425		430
Asp Ala Glu Ile Pro Leu Cys Lys Pro Val Asn Cys Gly Pro Pro Glu						
	435			440		445
Asp Leu Ala His Gly Phe Pro Asn Gly Phe Ser Phe Ile His Gly Gly						
	450			455		460
His Ile Gln Tyr Gln Cys Phe Pro Gly Tyr Lys Leu His Gly Asn Ser						
465		470		475		480
Ser Arg Arg Cys Leu Ser Asn Gly Ser Trp Ser Gly Ser Ser Pro Ser						

	485		490		495
Cys Leu Pro	Cys Arg Cys Ser Thr	Pro Val Ile Glu Tyr	Gly Thr Val		
500		505	510		
Asn Gly Thr Asp	Phe Asp Cys Gly Lys Ala Ala Arg	Ile Gln Cys Phe			
515		520	525		
Lys Gly Phe Lys Leu Leu Gly Leu Ser Glu Ile Thr Cys Glu Ala Asp					
530		535	540		
Gly Gln Trp Ser Ser Gly Phe Pro His Cys Glu His Thr Ser Cys Gly					
545		550	555		560
Ser Leu Pro Met Ile Pro Asn Ala Phe Ile Ser Glu Thr Ser Ser Trp					
	565		570		575
Lys Glu Asn Val Ile Thr Tyr Ser Cys Arg Ser Gly Tyr Val Ile Gln					
	580		585		590
Gly Ser Ser Asp Leu Ile Cys Thr Glu Lys Gly Val Trp Ser Gln Pro					
	595		600		605
Tyr Pro Val Cys Glu Pro Leu Ser Cys Gly Ser Pro Pro Ser Val Ala					
	610		615		620
Asn Ala Val Ala Thr Gly Glu Ala His Thr Tyr Glu Ser Glu Val Lys					
	625		630		635
Leu Arg Cys Leu Glu Gly Tyr Thr Met Asp Thr Asp Thr Asp Thr Phe					
	645		650		655
Thr Cys Gln Lys Asp Gly Arg Trp Phe Pro Glu Arg Ile Ser Cys Ser					
	660		665		670
Pro Lys Lys Cys Pro Leu Pro Glu Asn Ile Thr His Ile Leu Val His					
	675		680		685
Gly Asp Asp Phe Ser Val Asn Arg Gln Val Ser Val Ser Cys Ala Glu					
	690		695		700
Gly Tyr Thr Phe Glu Gly Val Asn Ile Ser Val Cys Gln Leu Asp Gly					
	705		710		715
Thr Trp Glu Pro Pro Phe Ser Asp Glu Ser Cys Ser Pro Val Ser Cys					
	725		730		735
Gly Lys Pro Glu Ser Pro Glu His Gly Phe Val Val Gly Ser Lys Tyr					

995 1000 1005
Cys Val Ala Pro Tyr Gln Cys Asp Cys Pro Pro Gly Trp Thr Gly Ser
1010 1015 1020
Arg Cys His Thr Ala Val Cys Gln Ser Pro Cys Leu Asn Gly Gly Lys
1025 1030 1035 1040
Cys Val Arg Pro Asn Arg Cys His Cys Leu Ser Ser Trp Thr Gly His
1045 1050 1055
Asn Cys Ser Arg Lys Arg Arg Thr Gly Phe
1060 1065

<210> 156
<211> 43
<212> PRT
<213> Homo sapiens

<400> 156
Ser Gln Lys Asp Leu Ile Gly Glu Ile Gln Ser Val Thr Cys Glu Glu
1 5 10 15
Val Tyr Ser Thr Pro Pro Pro Lys Lys Ile Phe Gly Ile Tyr Thr Ser
20 25 30
Arg Asn Arg Ala Asn Met Cys Gly Asn Ala Tyr
35 40

<210> 157
<211> 61
<212> PRT
<213> Homo sapiens

<400> 157
Val Asp Ser Pro Pro Thr Ser Pro Phe Val Ser Arg Pro Ile Ser Arg
1 5 10 15
Gln Glu Ser Gln Lys Asp Leu Ile Gly Glu Ile Gln Ser Val Thr Cys
20 25 30
Glu Glu Val Tyr Ser Thr Pro Pro Pro Lys Lys Ile Phe Gly Ile Tyr
35 40 45

Thr Ser Arg Asn Arg Ala Asn Met Cys Gly Asn Ala Tyr
 50 55 60

<210> 158

<211> 263

<212> PRT

<213> Homo sapiens

<400> 158

Leu Arg Lys Asp Thr Ala Gly Cys Phe Ser Gln Arg Arg Ser Arg Ala
 1 5 10 15

Met Leu Val Leu Arg Ser Ala Leu Thr Arg Ala Leu Ala Ser Arg Thr
 20 25 30

Leu Ala Pro Gln Met Cys Ser Ser Phe Ala Thr Gly Pro Arg Gln Tyr
 35 40 45

Asp Gly Ile Phe Tyr Glu Phe Arg Ser Tyr Tyr Leu Lys Pro Ser Lys
 50 55 60

Met Asn Glu Phe Leu Glu Asn Phe Glu Lys Asn Ala His Leu Arg Thr
 65 70 75 80

Ala His Ser Glu Leu Val Gly Tyr Trp Ser Val Glu Phe Gly Gly Arg
 85 90 95

Met Asn Thr Val Phe His Ile Trp Lys Tyr Asp Asn Phe Ala His Arg
 100 105 110

Thr Glu Val Arg Lys Ala Leu Ala Lys Asp Lys Glu Trp Gln Glu Gln
 115 120 125

Phe Leu Ile Pro Asn Leu Ala Leu Ile Asp Lys Gln Glu Ser Glu Ile
 130 135 140

Thr Tyr Leu Val Pro Trp Cys Lys Leu Glu Lys Pro Pro Lys Glu Gly
 145 150 155 160

Val Tyr Glu Leu Ala Thr Phe Gln Met Lys Pro Gly Gly Pro Ala Leu
 165 170 175

Trp Gly Asp Ala Phe Lys Arg Ala Val His Ala His Val Asn Leu Gly
 180 185 190

Tyr Thr Lys Leu Val Gly Val Phe His Thr Glu Tyr Gly Glu Leu Asn

195 200 205
 Arg Val His Val Leu Trp Trp Asn Glu Ser Ala Asp Ser Arg Ala Ala
 210 215 220
 Gly Arg His Lys Ser His Glu Asp Pro Arg Val Val Ala Ala Val Arg
 225 230 235 240
 Glu Ser Val Asn Tyr Leu Val Ser Gln Gln Asn Met Leu Leu Ile Pro
 245 250 255
 Ala Ser Phe Ser Pro Leu Lys
 260

<210> 159
 <211> 203
 <212> PRT
 <213> Homo sapiens

<400> 159
 Tyr Lys Pro Val Lys Trp Val Ile Ser Leu Thr Pro Leu Ser Gln Pro
 1 5 10 15
 Gly Pro Ser Ser Asn Ile Ile Gly Gln Ser Val Glu Glu Ala Ile Arg
 20 25 30
 Gly Val Phe Asp Ala Ser Leu Lys Met Ala Gly Phe Tyr Gly Leu Tyr
 35 40 45
 Thr Trp Leu Thr His Thr Met Phe Gly Ile Asn Ile Val Phe Ile Pro
 50 55 60
 Ser Ala Leu Ala Ala Ile Leu Gly Ala Val Pro Phe Leu Gly Thr Tyr
 65 70 75 80
 Trp Ala Ala Val Pro Ala Val Leu Asp Leu Trp Leu Thr Gln Gly Leu
 85 90 95
 Gly Cys Lys Ala Ile Leu Leu Leu Ile Phe His Leu Leu Pro Thr Tyr
 100 105 110
 Phe Val Asp Thr Ala Ile Tyr Ser Asp Ile Ser Gly Gly Gly His Pro
 115 120 125
 Tyr Leu Thr Gly Leu Ala Val Ala Gly Gly Ala Tyr Tyr Leu Gly Leu
 130 135 140

Glu Gly Ala Ile Ile Gly Pro Ile Leu Leu Cys Ile Leu Val Val Ala
 145 150 155 160
 Ser Asn Ile Tyr Ser Ala Met Leu Val Ser Pro Thr Asn Ser Val Pro
 165 170 175
 Thr Pro Asn Gln Thr Pro Trp Pro Ala Gln Pro Gln Arg Thr Phe Arg
 180 185 190
 Asp Ile Ser Glu Asp Leu Lys Ser Ser Val Gly
 195 200

<210> 160
 <211> 85
 <212> PRT
 <213> Homo sapiens

<400> 160
 Val Leu Thr Pro Gly Pro Glu Val Gly Cys Asn Cys Leu Trp Lys Ile
 1 5 10 15
 Leu Pro Leu Ser His Pro His Pro Leu Val Pro Ala Lys Asn Ala Gly
 20 25 30
 Asp Ser Ala Thr Met Leu Cys Val Pro Leu Asn Leu Cys Ala Ala Arg
 35 40 45
 Lys Ala Asp Val Gln Val Tyr Leu Asn Pro His Tyr Ser His Ala Pro
 50 55 60
 Asn Val His Gly His Phe Cys Asn Ser Ala His Ile Leu Thr Val Phe
 65 70 75 80
 Glu Lys Leu Ala Val
 85